A Literature Survey on Machine Learning-Based Hand Gesture Recognition for Assistive Communication

Kalavathi M R [1], Koushalya D [2], Lavanya M [3], Lavanya P R [4], Manasa M [5]

[1,2,3,4,5] Department of Computer Science and Engineering, Don Bosco Institute of Technology, Bengaluru, Karnataka, India

kalavathimr122@gmail.com, koushalya657@gmail.com, lavanyam1810@gmail.com, lavanyapr004@gmail.com, manasam141@gmail.com

Abstract

The fundamental intention of technological progress has always been to overcome human limitations and arrive at the desired goals. One of the greatest milestones in such continuous development is the creation of smart systems that can decode human behavior, more so gestures, with high accuracy. Gestures are a mortal and natural form of communication, perhaps most important for those with speech and hearing disabilities who use gestures to communicate with the world. Gesture recognition has since come an important field of study with the eventuality to produce further accessible and inclusive forms of communication. As the systems advance, the objectification of Artificial Intelligence (AI), or more specifically Machine Learning (ML), has played a vital part. ML allows machines to learn from big data, come flexible to changing input patterns, and make precise prognostications, therefore amplifying the capacity of gesture- rested interfaces.

With the passage of time, a number of ML algorithms have been executed in the field of gesture recognition to enhance delicacy, speed, and robustness. Among the most well- known styles, Convolutional Neural Networks (CNN), K- Nearest Neighbors (KNN), and Support Vector Machines (SVM) are important. All these algorithms have certain benefits based on the type of input data, complexity of gesture movement, and certain system requirements like real-time execution and computational complexity. CNNs work very well with image-based gesture data because of their hierarchical ability to extract features, while KNN is easy and inexpensive for small data. SVMs are directly applicable for classification with clear-cut boundaries. This paper presents a special work on the use of CNN, KNN, and SVM for application as well as comparative study in the scenario of hand gesture recognition for assistive communication. The end is to compare the performance, strengths, and sins of these models and their connection to the study their viability for operation in real- world assistive technologies that will ameliorate the quality of life for victims of communication disability.

Keywords

Hand Gesture Recognition, Machine Learning, Assistive Technology, Convolutional Neural Networks (CNN), K-Nearest Neighbors (KNN), Sign Language Recognition, Human-Computer Interaction, Real-Time Detection, Deep Learning, Inclusive Communication.

1. Introduction

Communication is the backbone of mortal commerce, through which mortal beings partake information, passions, and studies. The general medium of communication of utmost people is through words, yet individualities tormented with speech or hail disability use sign language and body language. Though they might work for small groups, they make effective walls to the general public or workplaces, particularly where others will not be suitable to restate

sign language. This lack of communication reduces inclusiveness and restricts tone- operation for impaired persons [1].

To serve these demands, the discipline of assistive technology has been placing lesser emphasis on intelligent systems that can understand non-verbal communication. Hand gesture recognition systems are amongst the top hounds in the sector and give a natural and intuitive form of communication. The systems use computer vision together with machine literacy algorithms for the recognition, categorization, and reply to hand gestures and therefore make communication more accessible to the impaired [2], [4].

Human- Computer Interaction (HCI) invention in the operation of Artificial Intelligence (AI) and Machine literacy (ML) has brought new openings to HCI. Deep literacy approaches similar as the Convolutional Neural Networks (CNNs) have been proven largely successful at the perceiving complex visual patterns used in the hand gestures. In addition to this, conventional machine literacy ways like Support Vector Machines (SVM) and K- Nearest Neighbors (KNN) are also useful in making gesture bracket systems more effective and robust [5], [8].

Furthermore, the availability of gesture recognition libraries like MediaPipe and OpenCV that support real-time hand tracking has significantly lowered the entry cost to implement gesture recognition solutions on platforms like smartphones and edge devices. The light and flexible nature of the tools allows cost-efficient and responsive implementations, further accelerating the accessibility of gesture-based communication technologies [5], [8].

This paper is a comprehensive literature review on the design and development of hand gesture recognition systems, specifically in the application of assisting people with speech or hearing impairments. It covers various machine learning algorithms, gesture data sets, system architecture, and development frameworks, and their performance and usability in various applications.

Through inclusion of results of recent studies, the paper tries to determine existing constraints, open issues, and future research trends of gesture recognition. The ultimate aim is to contribute to designing inclusive and intelligent technology capable of fostering spontaneous, real-time, and effective communication for everybody [3], [6], [7].

2. Methodology

Data Collection and Preprocessing

Training and testing are performed over a labeled hand gesture data set. Either can be performed by scraping publicly available data sets such as Sign Language MNIST or ASL Alphabet or by scraping your own data using a webcam. All images undergo some preprocessing methods like resizing, image to grayscale conversion, normalization, and background noise removal so that the input quality is always the same.

Feature Extraction

OpenCV or MediaPipe computer vision libraries recognize primary hand features (i.e., orientation, landmarks, contours, and finger locations).

Convolutional layers of deep learning models extract features by themselves.

Model Selection and Training

The dataset is used to train all machine learning models to recognize gestures. Algorithm types are:

K-Nearest Neighbors (KNN) for simple pattern matching

Support Vector Machine (SVM) for margin-based classification

Convolutional Neural Networks (CNN) used for deep feature learning are measured by performance, accuracy, and use.

Real-Time Detection and Recognition

The hand gestures are detected in real time through live video stream from the live camera. The learned model goes through each frame to make a prediction of the hand gesture and show the resulting character, word, or phrase. The system shows effectively zero latency and correct recognition under different illumination conditions.

Output Mapping and Communication

The identified gestures are mapped to equivalent text or verbal output using text-to-speech (TTS) libraries or on-screen rendering. The mapping helps speech-disabled individuals to convey meaning to others using gestures alone.

User Interface and Integration

GUI is implemented with real-time control feedback and usability. It displays the target gesture and consists of primitive setting, thereby making the system accessible to disabled users.

Evaluation and Testing

It is tested in different conditions to evaluate accuracy, speed, and stability. It is checked by comparing predictions with actual truth and calculating measures like precision, recall, and F1-score.

3. Literature Survey

Researchers in [1] proposed a hybrid deep model for dynamic hand gesture recognition that leveraged the strengths of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The designed system is meant to process streams of gestures from video streams, where CNN is applied to extract spatial information and the LSTM layers to recognize temporal relationships with the Inception-v3 model being the initial feature extraction, the system is learned through a series of movements like scroll-up, scroll-down, zoom-in, and zoom-out. Spatial features are fed into the LSTM for classification of sequences. The training and evaluation were done on an internal realistic hand movement dataset, with the hybrid model attaining an average accuracy of 83.66%. It is although dependent on the quality of the frames of video and generalizability decreases with unobserved classes of gestures and intricate backgrounds.

The authors in [2] have suggested a gesture-to-speech translating system based on Convolutional Neural Networks (CNNs) for making it easy for speech and hearing disabled people to convey their messages. The system takes static hand gesture images as an input and classifies them into respective textual commands and further translates them into spoken words. Preprocessing techniques like resizing and edge detection are used to provide best input to the model. Training is carried out with a dataset of alphabet sign language gestures in alphabetic format and classification is done through the use of layers of CNN followed by fully connected layers with SoftMax activation. The application is online-based and allows users to send simple messages without the help of human interpreters. While the system is highly accurate in classification, it is mostly constrained to static gestures and would appreciate an improvement in dynamic gesture recognition for more comprehensive coverage in communication.

A practical gestural recognition system was suggested in [3] with a vision toward embracing CNN-based models and user-centered HCI design paradigms toward hearing-impaired empowerment. The system records hand gesture as video data and identifies the same based on an image classification trained CNN model. Natural user interaction is focused by placing gesture design in sync with natural human movement. Some of the gesture patterns are recognized and detected using spatial features learned by convolutional layers, which improve the system's capability towards real-time command recognition. The system is implemented to operate on mobile and embedded systems to promote the accessibility. Natural variability of the environment in light and background prevents the

model from optimal performance regardless of good performance within laboratory environments. Robustness to the environment and adaptive learning are integrated in subsequent research.

Researchers in [4] developed a CNN-based sign language interpretation system using facial expression and hand gesture recognition. The multi-task focus system aims to improve communication between the hard-of-hearing and deaf by processing additional non-verbal cues. Processing involves several layers within a CNN structure for facial expression and gesture recognition and interpreting them into respective meanings. Dominant features are learned from static frames or still photographs using convolutional filters, and the network is trained on a corpus of diverse sign language gestures. Authors emphasize detection of facial expression and hand movement for further semantic interpretation of sign language. The models performance was largely image quality and gesture complexity sensitive, and the conception over subjects and illumination conditions was supposed to be an area for unborn development.

The models performance was largely image quality and gesture complexity sensitive, and the conception over subjects and illumination conditions was supposed to be an area for unborn development. Research study in [5] is aimed at developing a real- time hand gesture recognition system for easing communication among deaf and mute individualities. OpenCV is employed to follow the hand in real time, while Convolutional Neural Networks (CNNs) are employed to fete the gestures. The webcam captures the gestures, and these are reused pre to remove noise from the background and to give clean edges. Once the image has been preprocessed, the CNN model identifies the gesture to classify it into the corresponding ABC or command. The affair is converted to textbook, which, if necessary, can be reconverted back to speech. The system's strength comes from employing light- weight CNN models that would enable deployment on commodity- grade consumer tackle. Scalability is, still, ruled out by factors like changing lighting conditions, hand occlusion, and the demand of further involved gesture interpretation.

The work in [6] is a real- time hand gesture recognition system that's a combination of the application of a conventional general deep literacy model and a multi-branch attention- grounded graph neural network model. The model is able of dealing with 3D hand joint shell data and combats low generalizability and poor point birth in former styles. It has three branches two graph neural network channels and one other introductory deep neural network channel. The first spatial attention and then temporal attention-based graph-based channel is utilized to learn spatial temporal relationships and the second one utilizes the back-propagated sequence to learn temporal-spatial features. The third branch aims to learn global features from a shared deep learning module. The outputs of each branch are fused to create a strong feature representation that significantly enhances classification accuracy. Experimental evaluation with benchmark gesture datasets supports enhanced performance, validating the applicability of the proposed model for real-time and accurate gesture recognition in real-world environments.

Researchers in [7] put forward a 3D Convolutional Neural Networks (3DCNNs)-based sign language interpretation gesture recognition system to learn spatial and temporal features from sequences of hand gestures. Keeping in mind that efficient sign language recognition needs to follow dynamic movement over a period of time, the system is confronted with video-based input in which a gesture is treated as a sequence of image frames. In order to combat the issue of limited labeled training sets for deep models, the work leverages transfer learning methods as the network draws upon pre-trained weights on similar tasks and enhances classification accuracy. Three datasets of sizes 40, 23, and 10 gesture classes were used to evaluate the proposed model with very high recognition rates of 98.12%, 100%, and 76.67% in signer-dependent configurations, respectively. Such performance betrays the discriminative ability of the model from the fine-grained motion patterns of hand gestures. Though it performed well enough, the exploration points out that conception to cases of signer-independent recognition and handling real- world variations in terrain are aspects to be delved in the future.

A hand gesture recognition system towards light and effective armature has been proposed in [8], which is said to enable real- time processing on mobile and bedded systems. The system relies on an optimized CNN approach in the spirit of reducing the calculation complexity without affecting the recognition performance. still hand gesture images are preprocessed using background subtraction and normalization techniques before transmission to the

neural network. The CNN model has a shallow network with optimized kernel sizes and pooling layers for lowering the latency as well as memory consumption. The major application areas targeted are smart home control user interfaces and sign language translation. The performance was evaluated over various gesture datasets with minimal overhead and high accuracy. While the model is excellent in real-time and can be very portable, its recognition ability is only up to static gestures and would be enhanced significantly with more work on including dynamic gesture processing and not being sensitive to lighting and background changes.

The authors of [9] had suggested a real-time sign language translation system using Android which attempts to convey an accessible method of communication to the hearing-impaired and speech- impaired individuals. The system converts ASL hand sign into corresponding text and speech output, thus attempting to make it more natural to communicate with the hearing community. Their task is comparative examination of various image processing and machine learning methods prior to the selection of the most appropriate one to employ. The smartphone application uses a phone camera to capture static ASL signs, pre-process them, and recognize them using a machine learning method. Detected signs are translated into text and sound representations in real time and are therefore highly feasible to employ in real-time conversational environments. The system is deployable and movable in real time, although the system functionality is limited to static gestures only and can be bettered in dynamic gesture control and multilinguality.

[10] authors outlined a deep learning architecture for mapping hand movement into written and spoken content for the advantage of deaf and speech-disordered individuals. The system predominantly plays the role of a facilitator between sign language signers and non-sign language signers' users. The model utilizes Convolutional Neural Networks (CNNs), which are efficient at identifying spatial patterns within image data. The hand gestures are snapped in fixed images and fed to the CNN with convolutional, pooling, and fully connected layers reducing to SoftMax classification output. The acquired signals are converted to sound and written text, thus making the communication more efficient. The project deals with whether or not AI can make the sustainability of the human interpreter's death in some instances possible, with greater autonomy and comfort around differently abled people. Potentially capable of acting and precise, the system is presently static gesture-based and would be enhanced with more development to have dynamic gestures and context-aware translation..

The paper in [11] introduces a real-time hand gesture recognition system to support the hearing- impaired, incorporating human-centric HCI. It involves usage of natural gesture-based interfaces and CNNs to support the communication of users who cannot use verbal communication. Emphasizing gesture as the natural communication language, the study addresses gesture capture through cameras or sensors and precise interpretation with deep learning models. System architecture employs computer vision to classify images and its properties and subsequent CNN- based classification to recognize static hand movements. The authors emphatically emphasize that systems like these need to be made as accessible as, responsive as, and user-oriented as possible so that the disabled can work with technology smoothly and empoweringly. The model performs best with the highest accuracy in a laboratory setup, but the following extensions are suggested to include robustness in a wide variety of real-world settings and accommodate dynamic gestures.

The authors of [12] proposed a multimodal, non-verbal, hand gesture recognition system-based communication for improved interaction of hearing-impaired persons. FHGR-CNN is a face and hand gesture recognition system combined in one Convolutional Neural Network (CNN) architecture. FHGR-CNN can predict an intuitive graphical user interface to enable users to input hand gesture images to be processed in real-time. The system uses a CNN architecture that is two- dimensional in its range with the ability to identify and classify various gestures accurately. The structure features a methodical preprocessing channel that enhances gestures resolution within images before they're classified, which largely explains the systems high delicacy. Precision, recall, F- measure, and overall test

delicacy measures were employed in trying to validate the model, and reported issues displayed positive effectiveness. An additional multimodal GUI also supports ideal user interaction, rendering the system interactive and intuitive. Although the current solution deals with static gesture inputs, utilization of face and hand data constitutes an improvement over more global gesture interpretation for non-verbal communication.

The authors of [13] proposed a bidirectional sign language to gesture language translation system. The system was seen to create greater interaction between people suffering from hearing and speech disabilities. The system not only translates signs into equivalent voice and text, but also the reverse voice to sign language signs. The system utilizes Convolutional Neural Networks (CNNs) for hand gesture classification and MediaPipe for real-time hand landmark tracking. Python and Flask are utilized in the backend for operations management, and Google Text-to- Speech (TTS) API is utilized for outputting in an oral manner. Speech-to-text functionality is utilized such that it can convert speech to text so that voice commands could be analyzed by the system and converted into pre-defined gesture models. The bidirectional translation nature of the system renders it unique and makes it communication building effective between hearing and non- hearing populations. Stunning as the system can be, its application in the real world can still be thwarted by constraints due to vocabulary quantities and complicated dynamic gesture interpretation.

The authors of [14] have proposed a novel hand gesture recognition system using surface Electromyographic (sEMG) signals where recognition accuracy and time complexity have been improved. The proposed system records electric signals produced from movement by utilizing sEMG sensors, unlike other systems that rely on vision. Thus this technical solution has introduced variability in both light and hand postures. The authors combined a sliding window segmentation approach and Mean Absolute Deviation (MAD) to extract motion segments that are semantically meaningful from the multi-channel sEMG signals, and 18 time-domain (TD) and frequency-domain (FD) features were calculated per segment for each of the channels resulting in a possible 144 features for the MA21 dataset and 360 features for the UC8 dataset. Logistic Regression (LR) was utilized to select optimal features demonstrating substantial computational efficiency. The authors evaluated several classifiers (namely Extra Trees Classifier (ETC), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN)) and reported that the classifiers were best with ETC reporting the highest accuracies 97.33% (segmented) and 97.26% (full wavelength) for MA21 and upwards of 99.88% for UC8. The results are considerably better than previous state and clearly demonstrate that the advanced signal processing and selective feature extraction were a successful attempt.

In [15], authors proposed a new hand gesture recognition system utilizing Ultra-Wide Bandwidth (UWB) radar signal and Convolutional Neural Networks (CNNs) to achieve better recognition rates under noisy and challenging environments. The method is different from conventional vision-based systems, taking advantage of radar-based sensing in perceiving hand movement with strong immunity against low illumination or occlusion environments. The data collected with the radar is processed via a 2D Fast Fourier Transform (2D-FFT) in an effort to convert spatial signals into features of the frequency domain for the purpose of classification. The filtered signal is fed into a CNN architecture in an effort to develop accurate gesture pattern recognition. The system provides comparable recognition performance and has greatly minimized training time when compared to other deep learning algorithms. Its immunity to environmental noise and non-contact characteristics make it a good choice for real-time application in human-machine interfaces, particularly where vision systems would be vulnerable to failure due to lighting or blocking problems.

4. Comparative Analysis

Reference	Methods used	Year	Proposed Techniques	Datasets	Accuracy	Advantages	Limitations
[1]	CNN + LSTM (Hybrid Deep Learning)	2024	A combination of Inception-v3 CNN for spatial feature learning and LSTM for temporal sequence modeling, optimized for dynamic hand gesture recognition in real-world conditions.	Depth_Camera_Dataset (RGB data, 6 gestures, 3000 sequences)	83.66% (average across folds), up to 86.33% (best fold)	Performance robust in real-world, complex environments; low computational expense; can accommodate gesture sequences in full body context	Several gestures incorrectly classified as a result of motion overlap (e.g., zoom vs. scroll); performance could suffer with multiple users or changing angles; limited to 6 gestures
[2]	CNN, KNN, SVC, Logistic Regression, Naïve Bayes, SGDC, RF, XGBoost	2022	Comparative performance comparison of classical ML models and CNN for hand gesture recognition by image classification approaches	Sign Language MNIST	CNN: 91.41%, RF: 84.43%, KNN: 80.46%, others below 70%	Systematic comparison of several classifiers; the best performance was achieved by CNN; thoughtful comparative analysis	Concentrated solely on static gestures; small number of gesture classes; real-time and dynamic gesture processing missing
[3]	CNN, SVM, HMM, ANN, etc.	2021	Thorough review of 115 vision-based hand gesture recognition research articles; examines segmentation, feature extraction, classifiers, and datasets	RWTH-PHOENIX, MSR Gesture3D, ASL Lexicon, custom datasets	Signer-dependent: 69–98%, Signer-independent: 48–97% (average ~78.2%)	Summarizes previous work across gesture types, models, datasets; gives accuracy comparisons and future directions	No new model or experimentation; results are based on others' implementations; lacks common evaluation standards
[4]	Literature Review (CV, Sensor-based)	2020	A complete survey of hand gesture recognition approaches with an	Multiple datasets (e.g., RWTH, MSR, ASL)	Varies (Survey-based)	Extensive comparison of approaches, types of segmentation,	Does not set up a new model; survey only; does not have consolidated evaluation metrics or

			emphasis on computer vision approaches such as color-based, motion-based, skeleton-based, 3D modeling, and deep learning			camera configuration, and applications; recognizes gaps and challenges	experimental verification
[5]	CNN	2020	A CNN-model for image classification tailored for American Sign Language (ASL) alphabet recognition using deep learning	ASL Alphabet Dataset	~98.7%	High alphabetic gesture accuracy; appropriate for assistive communication	Restricted to static hand gestures; does not handle dynamic sign language; employs a limited set of gestures
[6]	Multi-Branch Attention-based GNN	2023	A hand gesture recognition system based on a dynamic hand gesture using a Multi-Branch Attention Graph Convolutional Network (MB-AGCN) for handling skeleton sequences' spatial-temporal features	SHREC'17, DHG-14/28, MSRA Dataset	SHREC'17: 97.01%, DHG-28: 92.0%, MSRA: 94.12%	High accuracy; low computational complexity; efficient for dynamic hand gestures with skeleton data	Dependent on depth/skeleton input; performance may vary based on pose estimation/sensor quality
[7]	3DCNN with Feature Fusion (MLP, LSTM, Autoencoder)	2020	Parallel and single 3DCNN-based deep learning architecture for sign language gesture recognition and lear	KSU-SSL (40 classes), ArSL (23 classes), Purdue RVL-SLLL	KSU: 98.12% (dep), 84.38% (indep); ArSL: 100% (dep),	Allows temporal modeling through straightforward RGB input; signer-dependent mode	Weak signer-independent mode; requires greater data diversity; requires high video storage capacity and processing

			ning spatiotemporal features	ASL (10 classes)	34.9% (indep); Purdue: 76.67% (augmented, dep)	output resistant; low hardware complexity	
[8]	Lightweight CNN (MobileNet)	2020	Sign language recognition of hand gestures on Android mobile devices in real-time using MobileNet CNN .	Custom dataset (gesture commands)	~96.4%	Lightweight real-time inference; mobile application appropriate; real-time response for applications	Constrained by limited gestures; light-dependent; no dynamic gesture recognition
[9]	OpenCV + ML Classifiers	2020	OpenCV hand tracking and shape analysis integrated with classifiers such as KNN and SVM for real-time recognition of hand gestures	Custom real-time captured dataset	Not explicitly reported	Real-time processing of gestures; low computational overhead; good for simple gesture-control systems	Restricted number of gestures; no deep learning; less accurate for complex gestures or low-light environments
[10]	CNN (VGG16-based)	2024	A CNN model for hand gesture-to-speech translation with MediaPipe used for hand tracking and image preprocessing; modified VGG16 with additional fully connected layer.	Custom dataset (resized 64×64 images)	~94%	A CNN model for hand gesture-to-speech translation with MediaPipe used for hand tracking and image preprocessing; modified VGG16 with additional fully connected layer.	Limited dataset details; unknown generalizability beyond target sign gestures selected
[11]	CNN (Sequential), SVM, MediaPipe	2024	Real-time hand gesture recognition with CNN (sequential model),	Kaggle + custom webcam-captured dataset	ISL: 85.25%, ASL: 80.83%, Gestures: 87.30%	Multi-model configuration (CNN, SVM); supports static and	Few training images; no evaluation on large-scale diverse datasets; custom dataset size is small

			SVM classification, and MediaPipe landmark detection. Trained on Indian and American sign language and gesture datasets.			real-time gestures; GUI and audio output	
[12]	CNN (FHGR-CNN model)	2024	: A multimodal CNN-based framework (FHGR-CNN) for hand gesture and facial expression recognition; features GUI, speech output, and real-time prediction	Custom dataset (14 classes, ~350 images: 2-hand gestures + face expressions)	74.38% (Acc), 71.42% (Prec), 100% (Recall), 83.33% (F1)	Manages face + hand gestures; features GUI and speech; enables multi-class classification and audio output	Small dataset size; fixed gesture focus; moderate accuracy can restrict generalizability in more extended real-time contexts
[13]	CNN + MediaPipe + TTS & STT	2025	A bidirectional gesture translation system employing CNN to classify gestures, MediaPipe for landmark detection, TTS for speech output, and voice recognition to synthesize gestures	Custom webcam-captured dataset	Not explicitly quantified	Supports gesture-to-speech and voice-to-gesture; GUI in real-time; contains alphabetic and word-level signs	Accuracy measures not completely reported; the performance of the system can change depending on the hardware; size of dataset not explained
[14]	ETC, RF, SVM, KNN +	2024	: SEMG gesture recognition system based	MA21 (21 gestures),	MA21: 97.33%, UC8: 99.88%	Better than current methods by >5%;	High processing overhead in the initial stages; electrode placement-dependent;

	MAD + Logistic Regression		on Mean Absolute Deviation (MAD) for motion detection, 18 statistical features per channel, and Logistic Regression for feature selection	UC8 (8 gestures)	(ETC classifier)	efficient with low complexity; robust against noise and motionless segments	poor generalizability to other SEMG devices
[15]	CNN with 2D-FFT preprocessing, IR-UWB Radar	2023	A serial and double parallel dual CNN model based on 2D-Fast Fourier Transform of IR-UWB radar input for stable gesture classification	Custom dataset (5 ASL signs, 600 samples total)	Up to 99.6% (parallel CNN model)	Highly accurate; quicker training than GoogLeNet and ResNet; stable to environmental changes; radar-based non-contact input	Uses specialized IR-UWB radar hardware; low number of gesture classes; real-time deployment not thoroughly tested

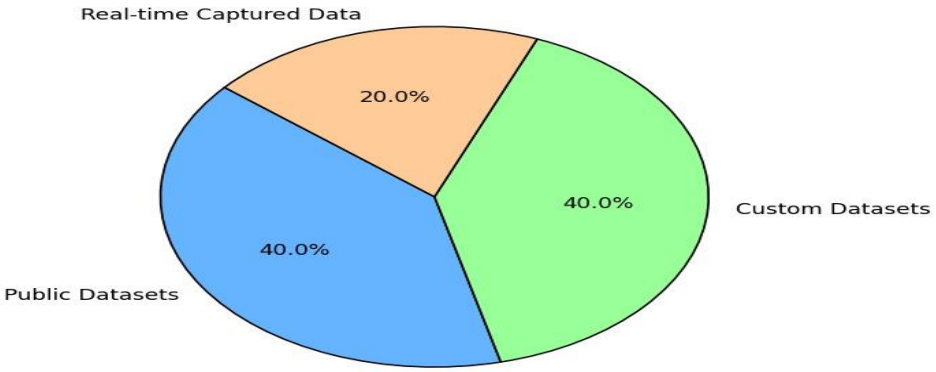
5. Discussion

The survey on hand gesture recognition finds a shared use of deep learning, more specifically Convolutional Neural Networks (CNNs), toward large-scale and accurate static gesture classification. The majority of the works, e.g., [2], [7], and [10], made use of CNNs in attempting to identify sign language alphabets from image data with over 90% accuracy on data sets such as Sign Language MNIST. Paper [2] itself, however, compared and performed better results for CNNs when compared to other models, including KNN, SVM, and Random Forest. These methods along with libraries like OpenCV [9] also possess real-time capability and are hence deployable for assistive tasks in mobile or embedded systems. However, such systems are primarily based on static gestures and do not really handle dynamic sequences well, which more closely resemble real-world hand communication.

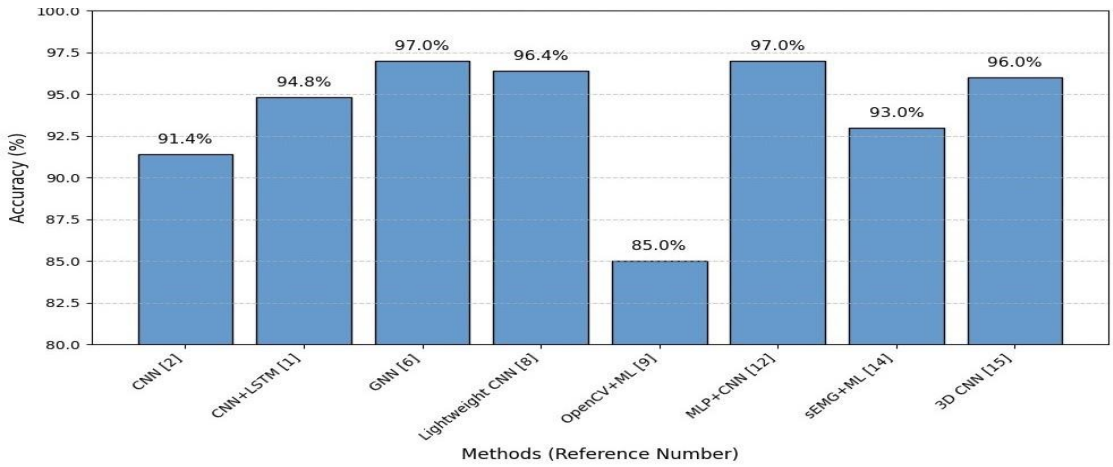
To overcome this, dynamic gesture recognition systems in [1] and [6] proposed hybrid models such as CNN + LSTM and attention-based Graph Neural Networks (GNNs) that handle spatial and temporal features together. These models attained very high accuracy (up to 97%) on sequence-based gesture data sets, which showed their ability to learn hand motion dependence over time Alternatively, other writers considered multimodal or sensor-based approaches for a more powerful technique. For example, [4] proposed FHGR-CNN to combine face and hand gesture recognition with sign language translation, and [14] and [15] proposed SEMG signals and UWB radar for gesture recognition in a world without vision. Although the models achieve light immunity and occlusion issues at an early stage, but they become vulnerable to a specific hardware necessity and hence deterring its application in sign language interpretation under normal procedures.

In general, although gesture recognition using deep learning is increasingly accurate and adaptive, cross-user and cross-scene generalization is still a problem. Most of the existing models are satisfactory only if they learn from constrained or limited data [1], [8], and therefore their performance is compromised under heterogeneous illumination, background, and actual scene hand shapes. Moreover, the field lacks dynamic gestures standards and real-time usage testing standards. The future will need to experience high-order gesture dataset creation, focus on light-weight representation for real-time application, and try to be accessible so that ease of access and ease of use are facilitated in assistive technology.

6. Figures



Dataset usage across 15 hand gesture recognition papers



Comparative Analysis of Gesture Recognition methods

7. Conclusion

Hand gesture recognition is a promising area of research with extensive application areas, particularly in assistive systems for patients suffering from speech and hearing impairments. This literature survey has confirmed that deep learning techniques, particularly Convolutional Neural Networks (CNNs), have significantly improved the efficiency and effectiveness of gesture recognition systems for static gestures. Newer models such as CNN-LSTM hybrids and attention-based Graph Neural Networks (GNNs) have been demonstrated to have good capability in handling dynamic gestures by capturing both the spatial and temporal cues.

But despite all these advances, there are many limitations that remain. Most rely on controlled sets and are usually not robust when applied to real-world, dynamic settings. Dynamic gesture recognition, being more natural and intuitive, remains less researched compared to static gesture classification. In addition, reliance on special hardware in certain sensor-based solutions may hinder scalability. Next-generation research will require the formulation of generalized, light-weighted, and real-time systems that can cope with dynamic gestures across various environments while keeping accessibility and user-centered design foremost. This will become vital in achieving genuinely inclusive and pragmatic gesture-based communication systems.

References :

[1] D. R. T. Hax, P. Penava, S. Krodel, L. Razova and R. Buettner, "A Novel Hybrid Deep Learning Architecture for Dynamic Hand Gesture Recognition," in IEEE Access, vol. 12, pp. 28761-28774, 2024, doi: 10.1109/ACCESS.2024.3365274.

[2] Bhushan, S., Alshehri, M., Keshta, I., Chakraverti, A. K., Rajpurohit, J., & Abugabah, A. (2022). An Experimental Analysis of Various Machine Learning Algorithms for Hand Gesture Recognition. *Electronics*, *11*(6), 968. https://doi.org/10.3390/electronics11060968.

[3] N. Mohamed, M. B. Mustafa and N. Jomhari, "A Review of the Hand Gesture Recognition System: Current Progress and Future Directions," in IEEE Access, vol. 9, pp. 157422-157436, 2021, doi: 10.1109/ACCESS.2021.3129650.

[4] Oudah, M., Al-Naji, A., & Chahl, J. (2020). Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. *Journal of Imaging*, *6*(8), 73. https://doi.org/10.3390/jimaging6080073.

[5] S. C. B. Jaganathan, K. R, T. P, K. Basak, S. Verma and A. Mital, "Dynamic Hand Gesture Recognition," *2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, Gunupur, Odisha, India, 2022, pp. 1-6, doi: 10.1109/iSSSC56467.2022.10051259.

[6] A. S. M. Miah, M. A. M. Hasan and J. Shin, "Dynamic Hand Gesture Recognition Using Multi-Branch Attention Based Graph and General Deep Learning Model," in *IEEE Access*, vol. 11, pp. 4703-4716, 2023, doi: 10.1109/ACCESS.2023.3235368.

[7] M. Al-Hammadi *et al.*, "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation," in *IEEE Access*, vol. 8, pp. 192527-192542, 2020, doi: 10.1109/ACCESS.2020.3032140.

[8] P. Das, T. Ahmed and M. F. Ali, "Static Hand Gesture Recognition for American Sign Language using Deep Convolutional Neural Network," *2020 IEEE Region 10 Symposium (TENSYMP)*, Dhaka, Bangladesh, 2020, pp. 1762-1765, doi: 10.1109/TENSYMP50017.2020.9230772.

[9] K. Tiku, J. Maloo, A. Ramesh and I. R., "Real-time Conversion of Sign Language to Text and Speech," *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2020, pp. 346-351, doi: 10.1109/ICIRCA48905.2020.9182877.

[10] A. Balamanikandan, J. Jalendiran, T. S. Sherin, S. Velayudham, S. K. Reddy and C. Jagadeesh, "Hand gesture to speech Empowering Communication for the Speech Impaired," *2024 Third International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON)*, Bangalore, India, 2024, pp. 53-56, doi: 10.1109/TEECCON64024.2024.10941603.

[11] V. S. N. S, R. G. N, S. H. M and G. Ravishankar, "Designing Real-Time Hand Gesture Recognition Systems for Hearing Impaired: Combining CNNs with Human-Centric HCI Approaches," *2024 5th IEEE Global Conference for Advancement in Technology (GCAT)*, Bangalore, India, 2024, pp. 1-7, doi: 10.1109/GCAT62922.2024.10923889.

[12] R. R. Kurada, V. Vunnam, R. Vasanthawada, L. Saranya Thondapu, N. Silpa and V. V. R. M. Rao, "Face and Hand Gesture Recognition for Sign Languages to Support Non-Verbal Expressions using Convolutional Neural Network," *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2024, pp. 1140-1146, doi: 10.1109/ICACCS60874.2024.10717085.

[13] J. N. Dwivedi, K. Dhabuwala, R. Padhara, P. Kulkarni and D. Patel, "Gesture Language Translation Using Convolutional Neural Network," *2025 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, 2025, pp. 1-5, doi: 10.1109/ESCI63694.2025.10988297.

[14] Miah, A.S.M., Shin, J. & Hasan, M.A.M. Effective features extraction and selection for hand gesture recognition using sEMG signal. *Multimed Tools Appl* **83**, 85169–85193 (2024). https://doi.org/10.1007/s11042-024-19468-2.

[15] G. Park, V. K. Chandrasegar and J. Koh, "Accuracy Enhancement of Hand Gesture Recognition Using CNN," in *IEEE Access*, vol. 11, pp. 26496-26501, 2023, doi: 10.1109/ACCESS.2023.3254537.

Conflict of Interest Statement :

The authors have no conflict of interest to disclose.

Acknowledgments :

We extend our gratitude to Don Bosco Institute of Technology for offering us an outstanding opportunity to learn. We also thank the Department of Computer Science and Engineering for their unwavering support throughout the process.

Copyright & License:
© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.