technologies

MDPI

Systematic Review

Gesture-Based Navigation of Smart Wheelchairs: A Review of Current Trends and Future Directions

Rakib Ahammed Diptho *, Safiul Haque Chowdhury, Md Abdullah Al Mamun, Md. Shakhawat Hosen, Md. Shamsur Rahman, Sarnali Basak and Md Abul Kalam Azad

Department of Computer Science and Engineering, Jahangirnagar University, Dhaka 1342, Bangladesh; 4169safiulhaque@juniv.edu (S.H.C.); mamun.stu2019@juniv.edu (M.A.A.M.); shakhawat.stu2019@juniv.edu (M.S.H.); drsrlabu19@gmail.com (M.S.R.); sarnali.cse@juniv.edu (S.B.); makazad@juniv.edu (M.A.K.A.)

* Correspondence: 3788diptho@juniv.edu

Abstract

Gesture recognition systems powered by artificial intelligence provide a promising solution for mobility and independence for individuals with physical disabilities. However, the deployment of such systems remains limited due to some challenges related to robustness, different user requirements, affordability for lower income people, and adaptation to low-resource environments. This study presents a systematic review of gesture-controlled intelligent wheelchair systems published recently. After searching academic databases, 600 studies were found. After removing duplicate and irrelevant studies and applying the inclusion and exclusion criteria, 72 of the most relevant studies were selected for detailed analysis. The review identifies three major approaches: vision-based methods, sensor-based techniques, and signal-based techniques utilizing electromyography (EMG) and inertial measurement units (IMU), and hybrid multimodal frameworks. A comparative study is conducted to analyze performance metrics, computational requirements, datasets, and validation strategies among these approaches. The findings identify several critical research gaps, including limited real-world testing, insufficient handling of pathological tremors, weak environmental robustness, and the lack of culturally aligned gesture vocabularies. The findings identify important design considerations and research directions for developing robust, affordable, and accessible intelligent wheelchair systems suitable for underserved people in low-resource environments.

Keywords: assistive technology; deep learning; gesture recognition; human–computer interaction; intelligent wheelchair; MediaPipe

Check for updates

Academic Editors: Daniele Giansanti and Hongying Meng

Received: 23 May 2026
Revised: 26 June 2026
Accepted: 11 July 2026
Published: 14 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. Background

Mobility is a crucial factor that enables individuals to live independently, participate in social activities, and maintain a good quality of life. However, individuals with physical disabilities often experience mobility-related impairments that create significant barriers to education, employment, and social inclusion. Globally, it is estimated that approximately 15% of the world's population lives with some form of disability, many of whom experience mobility limitations [1]. These challenges are particularly severe in lower-middle-income countries such as Bangladesh due to financial constraints, limited rehabilitation services, and restricted access to affordable assistive technologies.

To support independent mobility, wheelchairs are widely used as assistive devices. Manual wheelchairs provide basic mobility but require sufficient upper-extremity strength and coordination. Consequently, they are difficult to use for individuals with conditions such as spinal cord injury, muscular dystrophy, cerebral palsy, amyotrophic lateral sclerosis, and post-stroke hemiplegia. Although electric wheelchairs improve user independence, joystick-based interfaces remain challenging for individuals with severe upper-limb impairments. Furthermore, commercially available intelligent wheelchairs are often expensive, typically ranging from USD 1500 to over USD 10,000 [2], making them inaccessible to many users in resource-constrained environments.

To overcome these limitations, researchers have explored alternative human–machine interfaces (HMIs), including gesture recognition techniques. Gesture recognition systems interpret hand movements, head motions, facial expressions, or physiological signals to generate control commands for wheelchair navigation. In recent years, advances in artificial intelligence have significantly improved the performance of these systems. Deep learning architectures such as convolutional neural networks (CNNs) and MediaPipe-based vision approaches demonstrate strong performance in detecting both static and dynamic gestures. Similarly, YOLO-based frameworks improve environmental awareness through real-time object detection, while transformer-based architectures enhance spatial and temporal modeling capabilities. In addition, sensor-based systems using electromyography (EMG) and inertial measurement units (IMU) provide alternative control mechanisms for users whose hand movements are difficult to capture using vision-based approaches.

Although these technologies have shown promising performance, several challenges remain. Many proposed systems are evaluated only in controlled laboratory environments, limiting their generalizability in real-world conditions. Environmental robustness under varying lighting, crowded spaces, and outdoor conditions remains insufficiently addressed. Furthermore, many systems do not adequately consider pathological tremors or inconsistent motor patterns commonly observed in users with neurological disorders. Another limitation is the lack of culturally appropriate gesture vocabularies that reflect local user practices. Additionally, the integration of lightweight deep learning models suitable for edge deployment on low-cost embedded hardware remains limited [3,4].

Considering these existing limitations and research gaps, this study conducts a systematic review of gesture-based intelligent wheelchair systems published between 2022 and 2026. By analyzing recent research trends, methodologies, and performance characteristics, the study identifies important technological developments, existing research gaps, and future research directions for gesture-controlled intelligent wheelchair systems, particularly in the context of low-resource environments such as Bangladesh.

1.2. Article Search and Survey Methodology

This study follows the PRISMA framework to ensure transparency and reproducibility. Research published from January 2022 to February 2026 was included. The electronic database searches were executed between 1 February 2026 and 28 February 2026, and the review protocol (research questions, databases, search strings, and inclusion/exclusion criteria) was defined a priori before screening began. Databases searched included IEEE Xplore, Scopus, Web of Science, PubMed, SpringerLink, ScienceDirect, and Google Scholar. Additionally, relevant conference proceedings were reviewed in the areas of assistive technology, human–computer interaction and robotics; namely CHI, ASSETS, ICORR, IEEE SMC, and ROBIO.

A structured search strategy was developed using combinations of appropriate keywords related to the topic such as gesture recognition, hand gesture, wheelchair control, smart wheelchair, intelligent wheelchair, assistive technology, deep learning, computer

vision, MediaPipe, EMG, inertial sensors, Bangladesh, low-cost, and developing countries. Boolean operators were used to refine the search query and enhance the accuracy of the retrieval of relevant documents. To make the review reproducible, the exact Boolean query strings, the database fields searched, and the filters applied to each source are reported in Table 1. The core query string applied (with database-specific syntax adaptations) was: ("gesture" OR "hand gesture" OR "EMG" OR "IMU" OR "facial" OR "head movement") AND ("recognition" OR "control" OR "classification") AND ("wheelchair" OR "assistive mobility" OR "smart wheelchair" OR "intelligent wheelchair"). Searches were restricted to the Title/Abstract/Keywords fields (or each database's nearest equivalent) and limited to peer-reviewed journal articles and conference papers published between January 2022 and February 2026 in English. Because Google Scholar does not support strict field-limited queries and returns a large volume of low-precision results, a bounded protocol was adopted: the same core query was issued, results were sorted by relevance, and the first 200 records were screened; screening stopped after two consecutive result pages (20 records) yielded no new eligible study. Google Scholar was used primarily to capture grey literature and to verify that no major peer-reviewed study had been missed by the indexed databases; duplicates already retrieved from the indexed databases were removed during de-duplication. As a result of the initial search, 600 records were identified.

Table 1. Database-specific search strings, fields, filters, and records retrieved (searches executed February 2026).

Database	Search String (Core Query, Adapted to Syntax)	Fields/Filters	Records
IEEE Xplore	("gesture" OR "EMG" OR "IMU" OR "facial") AND ("recognition" OR "control") AND ("wheelchair" OR "assistive mobility")	Metadata; 2022–2026; Conf. + Journal	168
Scopus	TITLE-ABS-KEY(gesture OR emg OR imu) AND TITLE-ABS-KEY(wheelchair OR "assistive mobility")	Title/Abs/Key; 2022–2026; English	121
Web of Science	TS=((gesture OR EMG OR IMU) AND (wheelchair OR "smart wheelchair"))	Topic; 2022–2026; Article + Proc.	94
PubMed	(gesture[Title/Abstract] OR EMG[Title/Abstract]) AND (wheelchair[Title/Abstract])	Title/Abstract; 2022–2026	27
SpringerLink	(gesture OR EMG OR IMU) AND (wheelchair) AND (recognition OR control)	Title + Abstract; 2022–2026	58
ScienceDirect	(gesture OR EMG) AND (wheelchair) AND (recognition OR control)	Title/Abstract/Keywords; 2022–2026	49
Google Scholar	gesture (recognition OR control) wheelchair (EMG OR IMU OR vision)	Relevance-sorted; first 200 screened	83
Total records identified			600

Following the removal of 137 duplicate records, the titles and abstracts of the remaining 463 records were screened against inclusion and exclusion criteria established a priori. To reduce selection bias, study selection and data extraction were performed independently by two reviewers; disagreements were resolved through discussion, and any remaining conflicts were adjudicated by a third senior reviewer. Inter-reviewer agreement was substantial to almost perfect, with a Cohen's kappa of 0.81 at the title/abstract stage and 0.86 at the full-text stage.

The inclusion criteria were specified as follows. (IC1) The study presented original research on vision, sensor, or signal-based gesture/intention recognition that was either integrated into a wheelchair (or comparable powered mobility) platform or provided a clearly transferable technical contribution (algorithm, feature extraction pipeline, or dataset) applicable to wheelchair control. (IC2) For the purpose of this review, a "wheelchair system" was defined as a powered mobility platform driven by an explicit user intention interface, whether realised as a full hardware prototype, a simulated controller, or an

explicitly proposed control architecture. (IC3) The minimum acceptable level of assessment was the reporting of at least one quantitative performance measure (e.g., accuracy, F1 score, mAP, Cohen's kappa, error rate, or latency). (IC4) The study addressed, or was directly transferable to, the target user population (individuals with motor impairments or reduced mobility). (IC5) The work was peer reviewed, written in English, and published between January 2022 and February 2026. The corresponding exclusion criteria were: (EC1) absence of any quantitative evaluation; (EC2) review or survey articles without an additional technical contribution; (EC3) duplicate or superseded versions of the same work; (EC4) gesture recognition studies with no plausible mapping to wheelchair control and no assistive framing; and (EC5) studies whose full text could not be retrieved.

Two scope related clarifications are warranted. First, some included studies examine adjacent topics such as hand gesture recognition, sign language, biometric authentication, radar/Wi-Fi sensing, gaze interaction, and virtual reality rather than wheelchair systems directly. These studies were retained only when their algorithms, feature extraction methods, or evaluation frameworks were transferable to gesture-based wheelchair control and are explicitly identified as technical transferability evidence throughout the review. Second, studies involving non-disabled participants or generic datasets were included for their methodological contributions but were not treated as evidence of clinical readiness. Conclusions regarding real-world usability and robustness to pathological tremors are based solely on studies evaluating the target population. This distinction is reflected in the comparative analysis (Section 2) and the risk-of-bias assessment (Section 4).

Title and abstract screening excluded 312 records that did not meet the criteria, leaving 151 reports sought for full text assessment. During full text eligibility assessment, a further 79 reports were excluded for the following documented reasons: 22 reported no quantitative performance metrics (EC1), 31 concerned gesture recognition with no demonstrable relevance or transferability to wheelchair control (EC4), 6 were not available in English (EC5/IC5), 12 were review articles without an additional technical contribution (EC2), and 8 could not be retrieved in full text (EC5). This yielded a final set of 72 studies included in the qualitative synthesis. For all 72 included studies, data extraction was performed to obtain information regarding the publication metadata, research objectives, gesture recognition methodology, type(s) of sensor(s) used, machine learning/deep learning architecture(s) used, characteristics of the dataset, performance metric(s), integration strategy(ies), validation method(s), limitations of the study, and proposed direction(s) for future research.

Data from each of the papers selected for inclusion in this review were examined using thematic synthesis to identify common methodologies, comparative trends in performance, technological advancements, and persistent research gaps. The entire study selection process is illustrated in Figure 1 using a PRISMA flow diagram.

1.3. Contribution

This review provides a structured analysis of gesture-based intelligent wheelchair systems and offers a context-aware framework tailored for Bangladesh. The major contributions of this study are summarized as follows:

- A comprehensive review of vision-based, sensor-based, and hybrid gesture recognition approaches for intelligent wheelchair control.
- A comparative analysis of Artificial Intelligence (AI) and machine learning (ML) methodologies employed in recent literature.
- Identification of performance benchmarks and key research gaps in existing intelligent wheelchair systems.

- Proposal of a context-aware development framework considering economic, environmental, cultural, and healthcare constraints specific to Bangladesh.
- An integration strategy incorporating lightweight deep learning models suitable for edge deployment.
- Introduction of hybrid sensor-vision architectures to enhance system robustness and reliability.
- Emphasis on gesture compensation mechanisms to support users with motor impairments.
- Development of culturally relevant gesture vocabularies aligned with Bangladeshi social and communication conventions.

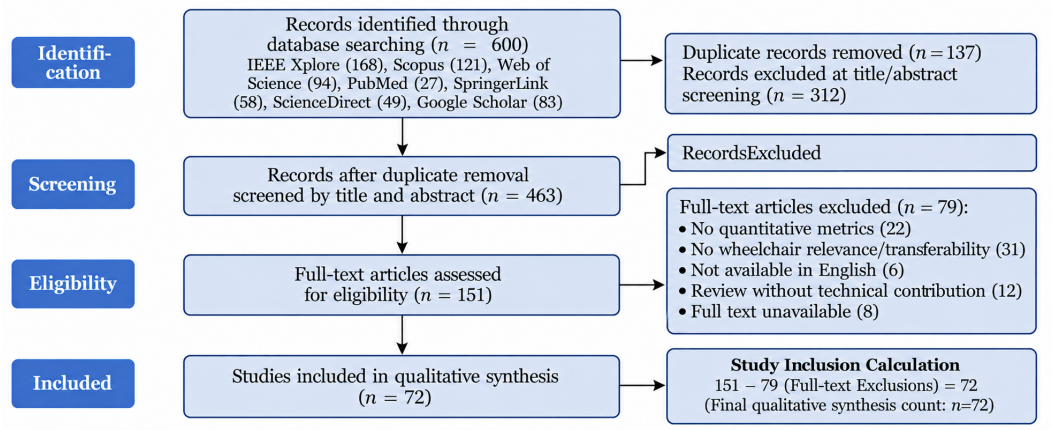


Figure 1. PRISMA 2020 flow diagram of the study selection process, regenerated with explicit record counts and documented exclusion reasons at each stage.

To avoid conflating evidence with design intent, we explicitly delineate the two types of contribution above. The first three items constitute the systematic-review evidence: they are conclusions derived directly from the synthesis of the 72 reviewed studies (comparative analysis, performance benchmarks, and identified research gaps). The remaining items constitute the author's design proposal: a context-aware development framework for Bangladesh that is informed by, but not directly evidenced by, the reviewed literature. Throughout the manuscript, claims drawn from reviewed evidence are kept distinct from elements of the proposed framework, and the framework itself is presented as a forward-looking synthesis in Section 4 rather than as a validated system.

1.4. Research Questions

To guide this review, the following research questions are formulated:

- Research Question 1 (RQ1): How have vision-based, sensor-based, and hybrid gesture recognition systems evolved for intelligent wheelchair control between 2022 and 2026?
- Research Question 2 (RQ2): What are the performance characteristics, deployment challenges, and contextual limitations of gesture-based intelligent wheelchairs in low-resource environments such as Bangladesh?

1.5. Organization of the Study

The remainder of this paper is organized as follows. Section 2 analyzes gesture-based control, covering sensor-based and vision-based gesture recognition systems. Section 3 examines signal-based control architectures and integration strategies, including mechanical, rule-based, biological, visual/optical, and audio-multimodal systems. Section 4 presents the overall discussion, including the proposed context-aware framework for Bangladesh and future research directions. Finally, Section 5 concludes the study.

2. Gesture-Based Control

This section analyzes studies of both camera-based (vision) and non-camera-based (sensor) input approaches to determine how data from each of these sources is processed using machine learning, deep learning, etc., for gesture recognition and wheelchair control.

Organizing principle and relationship between Sections 2 and 3. The reviewed literature is organized along the dimension that most strongly determines the recognition pipeline: the nature of the input modality. Section 2 covers gesture-based interfaces, in which the user performs an intentional, observable movement that is captured either by a camera (vision-based) or by a wearable or mounted motion sensor. Section 3 covers signal-based interfaces, in which the control command is decoded from a transduced physical or physiological signal that is not necessarily an externally observable gesture: mechanical pressure or tilt, surface electromyography, optical (visible light) modulation, or fused audio-sensor streams. We retain this two-level taxonomy because the engineering constraints (sensor placement, signal conditioning, real-time decoding, and calibration) differ markedly between the two families. We acknowledge, however, that the boundary is not strict: surface electromyography in particular can be framed both as a wearable gesture sensor (Section 2) and as a biological control signal (Section 3). To eliminate the duplication noted by the reviewers, each modality is described once in its primary category, cross-referenced where it recurs, and the categories are then reintegrated in a single cross-modality synthesis in Section 4.

Mathematical preliminaries. Several of the reviewed studies rely on a common set of time-domain feature descriptors and normalization operations. For convenience and to make the cited quantities explicit, the recurring formulations are defined here once and referred to thereafter; study-specific equations that are internal to a cited paper are referenced as such rather than reproduced. For a signal window $\{x_i\}_{i=1}^{N}$ with mean $\bar{x}$, the root mean square (RMS), mean absolute value (MAV), waveform length (WL), variance (VAR), standard deviation (SD), and zero crossings (ZC) are defined as:

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}, \qquad \mathrm{MAV} = \frac{1}{N}\sum_{i=1}^{N}|x_i|, \qquad \mathrm{WL} = \sum_{i=1}^{N-1}|x_{i+1} - x_i|, \tag{1}$$

$$\mathrm{VAR} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2, \qquad \mathrm{SD} = \sqrt{\mathrm{VAR}}, \qquad \mathrm{ZC} = \sum_{i=1}^{N-1} g(x_i, x_{i+1}), \tag{2}$$

where $g(x_i, x_{i+1}) = 1$ if $x_i\, x_{i+1} < 0$ and $|x_i - x_{i+1}| \geq \varepsilon$ (a noise threshold), and 0 otherwise. Feature vectors are commonly rescaled by min–max normalization, and geometric/landmark methods compute the Euclidean distance between two points $\mathbf{p}$ and $\mathbf{q}$ in $\mathbb{R}^n$:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \qquad d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{k=1}^{n}(p_k - q_k)^2}. \tag{3}$$

2.1. Sensor-Based Approaches

In this subsection, we analyze studies that employ sensor-based input devices, including accelerometers, gyroscopes, flex sensors, and Inertial Measurement Units (IMUs), for gesture recognition. The reviewed studies exclusively utilize non-camera-based approaches and rely on wearable or mounted sensors to capture and interpret user gestures for wheelchair control. A summary of these sensor-based wheelchair control systems is presented in Table 2.

Table 2. Sensor-based gesture wheelchair control systems.

Paper	Category	Sensor/Input	Dataset	Preprocessing	Model/Approach	Performance
Sensor—Deep Learning						
Vasconez et al. [5]	DL	Myo + G-Force	85 users	FIR, window	DQN + ANN	97.5–98.9%
Nogales & Benalcazar [6]	DL	Leap Motion	56 subj.	Window, stat. feat.	CNN + BiLSTM	99.99%
Wang et al. [7]	DL	IMU	20 users	EKF, Z-score	PTformer	Superior
Zhao et al. [8]	DL	MediaPipe LM	85 users	Resize	RNN/CNN	99.76%
Bao et al. [9]	DL	sEMG	8 pts.	BP filter	CNN/LSTM	72.95%
Zhang et al. [10]	DL	sEMG (Myo)	NinaPro	BW LP	CNN + Attn	96.47%
Fernandez et al. [11]	DL	Capacitive glove	12k	—	RF	97.11%
Kateb et al. [12]	DL	Textile capacitive	12 gest.	BW LP	kNN	100%
Kaur et al. [13]	DL	Gesture + FP	230 imgs	Aug.	YOLOv8 + CNN	98.8%
Sensor—Machine Learning						
Zhang et al. [14]	ML	Myo (8ch)	6 subj.	BW + PCA	DualTL	80.17%
Rusydi et al. [15]	ML	Flex + Gyro	20/6	Roll sep.	AHC	98.9%
Singh et al. [16]	ML	MPU6050	6 gest.	Auto feat.	MLP	84.67%
Anam et al. [17]	ML	Myo	5 subj.	Thresh.	kNN	96–100%
Sensor—Rule-Based						
Mahdin et al. [18]	Rule	Gyro + US + GPS	50 trials	Thresh.	Dir. map	∼64%
Calado et al. [19]	Rule	Glove + IMU	5000	Scale	Geom. model	92.1%
Chen et al. [20]	Rule	LeapMotion	20 subj.	Gaussian	GestureMoRo	0.08 err
Islam et al. [3]	Rule	MPU6050 + US	400	Offset	Thresh. map	95.5%
Rambabu et al. [21]	Rule	MEMS acc.	RT	Cal.	Tilt map	RT
Balaji et al. [22]	Rule	Head MPU6050	RT	LPF + MA	Axis map	RT
Sensor—Hybrid						
Gopal et al. [23]	Hybrid	12 EMG	NinaPro DB3	BW + RMS	Ensemble + CNN	Best F1

2.1.1. Sensor-Based Deep Learning Techniques

Deep learning is currently being used in an increasing number of ways to recognize gestures from the sensors that are attached to or integrated into our hands (for example, EMG, IMU, capacitive sensors, etc.) by providing both the ability to automatically extract features of interest and significantly better classification results than traditional machine learning methods. The current section provides a review of research using multiple deep learning architectures with different types of non-camera-based sensors as well as a comparison of the same research across three key areas: dataset properties; preprocessing techniques; and deep learning approaches.

Dataset: The datasets utilized vary considerably in number of subjects, gesture types, and sensor configurations. Vascenzez et al. [5] produced an enormous collection of data that contained 85 individuals who were all doing hand gestures (five static and six dynamic) and using a Myo armband (Thalmic Labs, Kitchener, ON, Canada) (8-channels at 200 Hz) and G-Force Sensor (OYMotion Technologies, Shanghai, China) (8-channels at 1000 Hz). The number of repetitions done by each individual was 180. Nogales and Benalcizar [6] built a Leap Motion (Leap Motion, Inc., San Francisco, CA, USA) database that consisted of five gestures being done by 56 people in total, which equaled to 84,000 observations at 70 Hz; thirty repetitions were done per person. Wang et al. [7] accumulated IMU data from twenty individuals who performed six different hand gestures; 2647 training samples were accumulated. Kateb et al. [12] used conductive elastic cord and twenty individuals to create a textile-based capacitive sensing system. Fernandez et al. [11] also made a dataset of 12,000 capacitance values from five sensors on a glove for the purpose of recognizing

sign language. Bao et al. [9] worked with eight chronic stroke patients who performed six hand gestures using sEMG signals from seven forearm muscles. Zhang et al. [10] utilized two datasets that were publicly available; one is the public NinaPro DB5 dataset (ten individuals, eighteen gestures), and the other is a private Myo dataset (ten individuals, six gestures). Kaur et al. [13] put together a fingerprint dataset of 230 images for the purpose of biometric authentication as well as collecting gesture data.

Preprocessing: Preprocessing strategies demonstrate common patterns while accommodating sensor-specific requirements. Vasconez et al. [5] did sliding window preprocessing (300-sample window, 40-sample stride), followed by finite impulse response (FIR) anti-aliasing filtering and time-domain feature extraction (RMS, SD, Energy, MAV, AE, where AE denotes the absolute envelope). Nogales and Benalcazar [6] did window splitting (windows of 20 samples, stride of 15 samples), followed by statistical feature extraction, including the myopulse percentage rate (MYOP), log detector (LD), waveform length (WL), enhanced waveform length (EWL), difference absolute standard deviation value (DASDV), and standard deviation (SD). Zhao et al. [8] used MediaPipe to locate the positions of 21 hand landmarks with 3D coordinates while reducing image size to 100×100 pixels. Wang et al. [7] performed z-score normalization and an extension of the Kalman filter for sensor fusion of nine axes, along with semi-automatic labeling. Bao et al. [9] performed zero drift correction and band-pass filtering of 20–450 Hz, as well as segmented data into segments of 102 ms with 51 ms increments. Zhang et al. [10] normalized the data and then performed a sliding window of 64 samples (320 ms) with a stride of 1 sample.

Approaches: The deep learning approaches demonstrate considerable diversity in architecture design. The first application of deep Q-Networks to recognize gestures from EMG-IMU was by Vasconez et al. [5]. They used DQNs to solve a problem of classifying gestures as a partially observable Markov decision process (MDP). Using the Myo armband they were able to achieve an accuracy rate of 97.50% for static gestures and 98.95% for dynamic gestures. Zhang et al. [10] designed a multi-attention CNN that had three types of attention, channel, spatial and temporal, which resulted in a 91.64% accuracy rate on the NinaPro DB5 database and a 96.47% accuracy rate on the Myo database; this was a 1.1% increase above the baseline. Zhao et al. [8] explored different architectures for recognizing hand movements from EMG signals, such as CNNs, RNNs, hybrid CNN-RNNs and Transformers. Zhao's team found that the RNN was the best architecture to use and it achieved an accuracy rate of 99.28%. The model also had a very low memory footprint of only 22 MB. Nogales and Benalcazar [6] looked at how well manual and automatic feature extraction worked for recognizing gestures from EMG signals. They manually extracted features using an ANN and a SVM, which resulted in accuracy rates of 92.94% and 91.37%, respectively. They automatically extracted features using a CNN and a BiLSTM, and they found that the BiLSTM-ANN method achieved an accuracy rate of 99.99% in only 27 ms. Wang et al. [7] created a new type of model called PTformer. PTformer uses two parallel transformers to capture the relationships across time and between different sensors. Kaur et al. [13] used a combination of YOLOv8 for gesture recognition (with an mAP50 of 98.8%), and contactless fingerprint biometric authentication (with an accuracy rate of 95%). Kateb et al. [12] created textile-based capacitive sensors that were capable of detecting 12 different gestures with an accuracy rate of 100% when using k-NN for classification. Fernandez et al. [11] used a random forest classifier to classify the data collected from capacitive sensors for recognizing American Sign Language, and their results showed an accuracy rate of 97.11%. Bao et al. [9] performed an exhaustive evaluation of 18 different deep learning models for recognizing hand movements from EMG signals. Bao's team found that frequency domain features provided the most accurate results for post-stroke patients, with an average accuracy rate of 72.95% per subject.

Challenges and Future Work: The study of Vasconez et al. [5] demonstrated the challenges of sensor variability. The authors further noted that the Myo armband performed better than the G-Force sensor. In addition, Wang et al. [7] found that the orientation of an inertial measurement unit (IMU) has a significant effect on the accuracy of stroke detection. Bao et al. [9] demonstrated that stroke patients experienced varying degrees of accuracy (85.84% to 69.43%) when using the same technology. Zhang et al. [10] were able to confirm that sEMG signals vary depending upon individual subjects. There is still a need to increase the amount of data available in order to make valid conclusions about the use of sEMG in stroke detection [6]. Additionally, real-time analysis can create trade-offs between accuracy and the time it takes to process [7]. Future directions for this research include developing end-to-end models [7]; investigating how to fuse multiple modalities such as EMG, kinematic, and/or acoustic sensors to improve stroke detection [10]; improving methods of transfer learning for clinical populations [9]; and compressing models to be deployable on embedded systems [8].

2.1.2. Sensor-Based Machine Learning Techniques

Machine learning approaches offer interpretable models and efficient computation suitable for embedded systems. This section analyzes studies employing traditional machine learning for gesture recognition using non-camera sensors.

Dataset: Zhang et al. [14] used the Myo armband (eight channels, 200 Hz) to collect sEMG data from six participants as they performed five static hand gestures (each of which was held for thirty seconds). Rusydi et al. [15] created a dataset utilizing two flex sensors and a single gyroscope with twenty test participants (with a total of 450 samples per gesture) and six test participants on a 58.8-m path. Singh et al. [16] utilized an MPU-6050 IMU (TDK InvenSense, San Jose, CA, USA) (six axes) to collect EMG signals for the performance of six different hand movements, collecting 256 data lines per gesture (or 256 points each) for edge computing applications. Anam et al. [17] collected EMG signals from five participants who were asked to perform five movements for thirty seconds each; this resulted in roughly 3000 data points per channel per movement.

Preprocessing: Zhang et al. [14] used a fourth-order Butterworth bandpass filter (30–70 Hz) and low-pass filter (60 Hz) with a 1-s sliding window (50% overlap) to extract seven time-domain and three frequency-domain features that were calibrated based on the subject mean. Rusydi et al. [15] separated their data into two algorithms based on the gyroscopic roll state of the user, utilizing Euclidean distance as a measure of similarity when applying AHC. Singh et al. [16] employed NanoEdgeAI for the automated extraction of features from 256×256 data buffers with both statistical analysis and inference functions which checked the data three times prior to classification. Anam et al. [17] normalized data by removing negative signals through threshold-based normalization and computed MAV, VAR, RMS and ZC features using a moving window (100-data-point shift). The output was then converted to ASCII for use in serial communication.

Approaches: Zhang et al. [14] proposed a method of dual-layer transfer learning (DualTL) that relies on the weak cross-user correlations as well as the strong within-user consistency. The first layer of DualTL identifies candidate solutions for the new user using the 5 nearest neighbors ($K_1 = 5$) with confidence threshold (μ) at 0.4. The second layer optimizes the candidate solution by maximizing both confidence and distribution divergence with λ at 0.5. Finally, the last layer of DualTL identifies the final classification solution for the remaining instances using a single nearest neighbor ($K_2 = 1$). On average DualTL was able to achieve an accuracy of 80.17%. This was a significant improvement over the other methods such as SMO which had an accuracy of 41.50%, KNNs at 41.05%, and RF at 44.43%. Similarly, DualTL had a better performance than TCA with accuracy

of 30.59% and JDA with accuracy of 30.84%. Rusydi et al. [15] utilized Agglomerative Hierarchical Clustering (AHC) with an average linkage to identify the five different types of wrist rotation; AHC was able to achieve a 100% in terms of training accuracy and a 98.88% in terms of testing accuracy in a real-world application. Singh et al. [16] developed a Multi-Layer Perceptron (MLP) using NanoEdgeAI on Nucleo-F401RE (STMicroelectronics, Geneva, Switzerland). In this research the MLP was able to achieve a 84.67% balanced accuracy with low-resource usage (3.2 KB RAM and 11.1 KB Flash). Anam et al. [17] performed a comprehensive k-NN analysis using multiple combinations of features and k values ($k = 3, 5, 9, 11$). It was found that when $k = 3$ the MAV + RMS feature combination was able to achieve a 100% accuracy, while the VAR feature combination was able to degrade the performance.

Challenges and Future Work: Zhang et al. [14] reported that the most important barrier to independence was the dual TL system itself with an accuracy of 80.17%, a value significantly lower than that required by natural HCI. Rusydi et al. [15] found that limited gesture sets and user adaptability were significant barriers. Singh et al. [16] indicated that while there is some promise for deploying this type of system at the edge due to the potential for high accuracy (84.67%), it will likely come at the cost of accuracy and may also limit the number of features that are included in the model, limiting its overall effectiveness. Anam et al. [17] reported that the inclusion of VAR consistently degrades performance. Real-world testing also demonstrated that the number of errors made in evaluating user's movements was higher when compared to their results from the offline evaluation. Therefore, future work should include advanced transfer learning [14], gestures that change dynamically with intuitive mappings [15], model compression to support edge deployment [16], and adaptive learning strategies [17].

2.1.3. Sensor-Based Rule-Based Techniques

Rule-based techniques provide a straightforward and easily understandable solution for creating wheelchair control systems based upon the application of rules (logical thresholds) applied to specific sensor readings to define specific wheelchair commands. The advantage of rule-based techniques is that they do not require any training data and, therefore, can be implemented on very low-cost microprocessors.

Dataset: A dataset was developed by Mahdin et al. [18], using gyro sensor, ultrasonic sensors, and GPS; this dataset has been tested in 50 trials for each of the object classes under the same conditions of illumination as well as under different conditions of illumination. Calado et al. [19] developed an Italian Sign Language Dataset, where five subjects performed ten Italian Sign Language words 100 times each (in total 5000 samples) in a 30–70 split training/testing set. Chen et al. [20] have conducted research with a LeapMotion sensor; twenty volunteers were measured using their LeapMotion sensors to measure palm distance, pitch and roll angle in 50 experiments. Islam et al. [3] collected the MPU6050 gyroscope data from 400 trials of four directional gestures plus stop (100 each); they achieved a success rate of 95.5%. Rambabu et al. [21] have also developed a system that utilizes MEMS accelerometer, DHT11 (Aosong Electronics, Guangzhou, China) and pulse sensor for health monitoring. Balaji et al. [22] developed a system utilizing head-mounted MPU6050 and calibration data to distinguish intentional from unintentional movements.

Preprocessing: Calado et al. [19] removed outliers by applying 99.5th and 0.5th quantiles, scaled flex sensor values to a $[-1, 1]$ square (scaled as defined in their work), and scaled IMU values to $[-1, 1]$ cube. The uniform linear interpolation of the sampled data allowed for a point-wise comparison of the two. Chen et al. [20] used Gaussian filter ($\sigma = 15$, window = 91) to remove palm jitter; however, they also experimented with an exponentially weighted moving average ($\alpha = 0.08$) and determined that it caused unaccept-

able delays. The depth information obtained through binocular vision was used to map the vertical distance to linear velocity (per their depth-to-velocity formulation) with thresholds in the range of 200 mm and 400–600 mm. Islam et al. [3] implemented gyroscopic calibration via the collection of N samples to determine offsets (per their calibration procedure), compared the calibrated data to thresholds with a delay of 100 ms between commands. Balaji et al. [22] utilized an adaptive filter that included low-pass filters and moving averages to differentiate between intended head movements; they communicated using ESP-NOW (Espressif Systems, Shanghai, China) for the purpose of achieving low latency communication.

Approaches: The geometric model-based algorithms μC^1 and μC^2 proposed by Calado et al. [19] are based on Clifford Algebra. These algorithms are used in combination with a Pointwise Euclidean Distance (pwD), and a Pointwise Multivector Distance (pwμVD) for μC^1. In addition to pwD and pwμVD, μC^2 includes Pointwise Multivector Euclidean Distance (pwμED) and Second-Order Multivector Distance (pwμ^2ED). The results of these two algorithms were an overall accuracy of 92.10% (with User Delay: 99.77% and User Intention: 91.56%), which took 11.43 milliseconds to recognize. A GestureMoRo algorithm was created by Chen et al. [20], which mapped the vertical palm distance to the linear velocity of movement and had five different angular velocity zones that corresponded to the orientation of the hand (based upon the roll angle). The average delay for this algorithm was approximately 168.06 ms and had an error rate of 0.08%. All twenty volunteers stated that this algorithm was either "easy" or "very easy". An ultrasonic sensor-based hand glove system with a threshold-based approach and a safety override using ultrasonic sensors was developed by Islam et al. [3]. This system utilized two complementary gesture-recognition algorithms and object detection using the YOLOv8 algorithm, which had a precision of 91.5%. Rambabu et al. [21] developed a tilt-based MEMS accelerometer control with RF transmission and ThingSpeak (MathWorks, Natick, MA, USA) health monitoring. A head-mounted MPU6050 with X/Y/Z threshold mapping and ESP-NOW communication was developed by Balaji et al. [22]. This system featured an IR obstacle detector and allowed users to customize their own locations.

Challenges and Future Work: Calado et al. [19] observed that μC^2 has slightly lower accuracy than deep learning techniques (BiLSTM-93.81%, GRU-93.24%). Chen et al. [20] concluded that Gaussian smoothing causes a delay in the system and that it is dependent upon the user having a stable Wi-Fi connection. Islam et al. [3] recognized that accuracy will be reduced when the system is used by someone making rapid movements and that consistency of the placement of sensors is required to monitor the health of the user. Rambabu et al. [21] also indicated that gesture control would not be effective for people who have difficulty using their hands. Balaji et al. [22] agreed that there is a need for greater distinction between intentional and unintentional head movement in the future of this technology. Mahdin et al. [18] stated that the greatest obstacle to developing a computer vision-based detection method is its sensitivity to light. Therefore, the area of focus for future research should include integrating machine learning into the recognition process to create an adaptive system [3]; extending the range of detection through the use of thermal or depth cameras [18]; implementing alternative Bluetooth-based backup systems [3]; creating an energy-aware scheduling methodology [21]; exploring the development of higher-order geometric models [19]; and incorporating Kalman filtering and sensor fusion [22].

2.1.4. Sensor-Based Hybrid Techniques

Hybrid approaches combine multiple machine learning and deep learning techniques to leverage complementary strengths, particularly valuable for challenging applications such as assistive robotics for amputees.

Dataset: Gopal et al. [23] used Nina Pro Database 3 (DB3) using four transradial amputees as subjects, which were representative of a variety in the type of amputation that the individuals experienced: Subject 1 (50% forearm was left intact), Subject 2 (70%), Subject 3 (30%), and Subject 9 (90%). Twelve EMG electrodes were located in an area surrounding the amputation site to capture data at 2 kHz. A total of ten gestures were taken from the original 50 gestures; these included finger movement, rotation, force application and grasp.

Preprocessing: Filters: Low Pass Filter (LPF): 1 Hz, 5 Hz, 10 Hz, 50 Hz; Band Pass Filter (BPF): 50–450 Hz (optimal was 1 Hz LPF). Rectification was used to obtain the signal envelope. A sliding window was used for the analysis, where the window size was one of five options: 50 ms, 100 ms, 150 ms, 200 ms, 250 ms (each window overlapped by 50%). Time Domain Features: RMS (Equation (1)), VAR (Equation (2)), EMAV (an enhanced mean-absolute-value variant) with $p = 0.75$ or $p = 0.50$, WL (Equation (1)). Five classifiers were tested in five configurations: Q1 (RMS), Q2 (EMAV), Q3 (WL), Q4 (VAR), Q5 (RMS + EMAV + WL + VAR). Cross validation with ten folds was employed.

Approaches: Gopal et al. [23] ran six feature-based methods in MATLAB (Tree, KNNs, SVM, LDA, Ensemble, ANN), and one non-feature-based method using a CNN with two convolutional layers, maxpool, dropout, batch normalization, and two fully connected layers in TensorFlow. The ensemble method produced the highest median accuracy, but the deep-learning methods (ANN and CNN) were able to produce higher F1-scores which indicates that these methods are better at predicting classes. The subject with the largest performance gap was Subject 3 (who had 30% of their forearm remaining) who used the deep-learning approach on their data when they had limited forearm data. Feature analysis with an Ensemble classifier (using 200 ms windows) demonstrated that combining the features from Q5 produced the highest accuracy across all subjects. The combined feature set (Q5) resulted in high accuracies of: Subject 1 (63.13%), Subject 2 (65.57%), Subject 3 (75.37%) and Subject 9 (70.70%). RMS (Q1) was the feature that worked best across all of the subjects.

Challenges and Future Work: Gopal et al. [23] identified primary limitations as requirements for large datasets; NinaPro DB3 contains few subjects, restricting generalization. Significant performance variability across amputation conditions (30–90% remaining forearm) highlights need for personalized models. Although deep learning algorithms have shown improved ability to generalize beyond training data, they require significant amounts of computation. Average accuracy, the "accuracy trap," may still mask poor generalization in individual classes. In addition, future research will need to be directed toward creating larger, more diverse databases of amputees; a detailed computational analysis to determine if there are real-time speed requirements; and the use of transfer learning using intact subject populations that also maintain performance with the amputee population.

2.2. Vision-Based Approaches

Vision-based gesture recognition provides an alternative approach for recognizing user gestures by utilizing cameras to detect hand and body movements instead of requiring users to wear sensors. This technology enables a more natural, contactless, and non-invasive method for wheelchair control. In this section, existing literature on vision-based gesture recognition methods is reviewed, covering a variety of camera-based approaches. These methods are categorized according to gesture type, including static gestures (hand poses) and dynamic gestures (movement trajectories). For clarity, we adopt the following definitions throughout this review. A static gesture is a posture defined by the spatial configuration of the hand (or face/body) at a single instant, for example a fixed hand shape or finger arrangement, and is recognized from one image frame without temporal information. A dynamic gesture is a gesture defined by motion over time, for example a

trajectory, velocity profile, or sequence of poses, and, therefore, requires a temporal window of frames (a video clip or skeleton sequence) for recognition. Equivalently, static gesture recognition is a spatial classification problem over a single frame, whereas dynamic gesture recognition is a spatio-temporal sequence modeling problem. Furthermore, each gesture category is analyzed based on the methodologies employed, namely deep learning (DL), machine learning (ML), and rule-based techniques. A comparative summary of these vision-based gesture recognition systems is presented in Table 3.

Table 3. Vision-based gesture wheelchair control systems (static and dynamic). A* denotes the A-star path-planning algorithm.

Paper	Category	Sensor/Input	Dataset	Preprocessing	Model/Approach	Performance
Vision—Static—Deep Learning						
Gadekallu et al. [24]	Static-DL	RGB images	20,000 images	Augmentation	CNN + HHO	100%
Bhushan et al. [25]	Static-DL	Sign MNIST	24,000 images	Feature selection	CNN	91.4%
Sadi et al. [26]	Static-DL	Custom RGB	900 images	YCrCb segm.	2D CNN	97.1%
Sahoo et al. [27]	Static-DL	MU + HUST-ASL	2515/5440	Depth threshold	AlexNet + VGG16	90–98%
Zhou & Chen [28]	Static-DL	OUHANDS	3000 images	DRN + ASPP	Dual CNN	91.2%
Dang et al. [29]	Static-DL	HANDS/SHAPE	12 k–30 k	HRNet	MobileNetV2/CNN	94–98%
Padhi & Das [30]	Static-DL	HaGRID	1900 images	MediaPipe	DenseNet201	97.6%
Mohamed et al. [31]	Static-DL	Custom	105,600 images	Gray + flip	CNN	∼99%
Wu et al. [32]	Static-DL	Custom	2850 images	Mosaic + HSV	YOLOv5s	96.8%
Jafari & Basu [33]	Static-DL	Multi-dataset	240–87 k	Resize 32 × 32	2DPSTPP-Net	98–100%
Kumar et al. [34]	Static-DL	Custom	60,000 images	Flip + blur	2-layer CNN	93%
Tran & Nguyen [35]	Static-DL	Custom	32,400 images	Not specified	ResNet18	99–100%
Vision—Static—Machine Learning						
Nivash et al. [36]	Static-ML	RGB + Face	24,000 images	HSV segm., mask	CNN + FaceNet	98.9/97.2%
Khaksar et al. [37]	Static-ML	MediaPipe (21 LM)	80 imgs, 8 gest.	LM scaling, angles	SVM/ANN/LR	96.3%
Vision—Static—Rule-Based						
Huda et al. [38]	Static-Rule	RGB Camera	20 smp/gest.	LM normalization	Distance thresh.	99.17%
Huda et al. [39]	Static-Rule	RGB Camera	700 samples	Dynamic ranges	Range logic	98.14%
Dragoi et al. [40]	Static-Rule	Laptop cam	31 users	Geom. finger logic	Rule mapping	86.7%
Ritu et al. [41]	Static-Rule	Webcam (face)	100 trials	MAR + nose zone	Facial logic	0.5–1.2 s
Vision—Static—Hybrid						
D'Souza et al. [42]	Static-Hybrid	RGB Camera	Gesture frames	Segmentation	CNN control	Intuitive
Mahdin et al. [18]	Static-Hybrid	Gesture + Obstacle	MS-COCO + gest.	Resize + norm.	CNN + Detection	Safe nav.
Abiraj et al. [43]	Static-Hybrid	Face + Gesture	Multimodal	Edge + ROI	Multimodal fusion	Robust nav.
He et al. [44]	Static-Hybrid	RGB dyn. gesture	Video dataset	GMM + PSO	Tracking + DL	High acc.
Vision—Dynamic—Deep Learning						
Gonzalez Leon et al. [45]	Dynamic-DL	RGB + Depth	762 sequences	Resize 120×160	3D CNN	99.48%
Peral et al. [46]	Dynamic-DL	IPN Hand	4218 instances	LM dist/time	Dense Network	87.5%
Riaz et al. [47]	Dynamic-DL	20BN-Jester	30,000 videos	30-frame unif.	3D-CNN + LSTM	97%
Nguyen et al. [48]	Dynamic-DL	IPN Hand	4000+	TD-Net features	TD-Net	84.98%
Miah et al. [49]	Dynamic-DL	MSRA/DHG/SHREC	2800–76,500	Graph embedding	Attn Graph DL	97.01%
Mohammed et al. [50]	Dynamic-DL	SHREC/DHG	2800+	Temporal norm	MMEGRN ensemble	96.43%
Narayanan et al. [51]	Dynamic-DL	Radar	2000+	5D point cloud	Deformable DETR	60.89 mAP
Wang et al. [52]	Dynamic-DL	FMCW radar	7000+	RTM/DTM/ATM	F-RCNN + GAN	80.8 mAP
Bremer et al. [53]	Dynamic-DL	VR + EEG	166 k samples	z-score	Transformer	78%
Vision—Dynamic—Machine Learning						
Bandara et al. [54]	Dynamic-ML	Leap Motion	25 users, 12 g.	3D grid partition	Two-stage NN	Kappa 0.984
Vision—Dynamic—Rule-Based						
Cui et al. [55]	Dynamic-Rule	Gest. + Voice + Head	10 users, 50 reps	Continuous cap.	Rule mapping	98.2%
Chamalsha et al. [56]	Dynamic-Rule	RGB Webcam	560 samples	Min-max scaling	Landmark logic	91.55%
Vision—Dynamic—Hybrid						
Sirisati et al. [57]	Dynamic-Hybrid	RGB video	20BN-Jester sub.	Convex hull + cont.	YOLO + LSTM	YOLO best
Meghna et al. [58]	Dynamic-Hybrid	Camera + LiDAR	Simulation	Grid conversion	YOLOv4 + A*	100% detect

2.2.1. Static Vision

Static gesture recognition focuses on identifying hand poses at a specific moment, characterized by spatial arrangement of fingers, hand shape, and orientation. Systems analyze individual image frames to classify predefined hand poses for wheelchair control commands.

Deep Learning Techniques

Deep learning has revolutionized vision-based static hand gesture recognition by enabling automatic feature extraction from raw images, eliminating manual feature engineering.

Dataset: The Gadekallu et al. [24] study utilized a Kaggle dataset containing over 20,000 images that were divided into ten different hand gestures (each gesture having 2000 images). Bhushan et al. [25], in turn, used a dataset called the "Sign Language MNIST" with over 24,000 images representing twenty different hand gestures. Sadi et al. [26] constructed their own custom dataset for use in testing the ability of wheelchair users to control their wheelchairs using hand gestures; this dataset contained 750 training images, 150 validation images, and was tested against the hand gestures of 700 elderly users. Sahoo et al. [27] also constructed two other datasets; one was the MU dataset, which consisted of 2515 images that represented thirty-six different hand gestures performed by five different individuals. The second was the HUST-ASL dataset, which had 5440 images and represented thirty-four different hand gestures performed by ten different individuals. Zhou and Chen [28] used a dataset called the "OUHANDS" dataset that included 3000 images of ten different hand gestures performed by twenty-three different individuals, as well as the ground truth for segmenting each image. Dang et al. [29] evaluated the performance of three different datasets, namely, the HANDS dataset that has 12,000 images of fifteen different hand gestures, the OUHANDS dataset with 3000 images of ten different hand gestures, and the SHAPE dataset that contains over 30,000 images of thirty-two different hand gestures. Padhi and Das [30] used a dataset called the "HaGRID subsample" that contained 1900 images representing eighteen different hand gestures. Mohamed et al. [31] generated 105,600 grayscale images of forty-four different hand gestures. Wu et al. [32] collected 2850 images of fourteen different hand gestures. Jafari and Basu [33] evaluated their system's performance on six different datasets that include ASL with 87,000 images, HGD with 12,064 images, etc. Bhavarthi et al. [59] constructed their own custom dataset to test the ability of wheelchair users to perform a series of five gestures. Madaan et al. [60] used the same Kaggle dataset used in the previous study but reduced the number of images from 20,000 to 2604 images and further divided the dataset into five different gestures. Awaluddin et al. [61] constructed a green-screen dataset that they duplicated thirty times and replaced the background of the images with the same color. Kumar et al. [34] constructed an additional 60,000-image dataset to be used to train their system to recognize five different hand gestures under various light conditions. Tran and Nguyen [35] constructed a dataset consisting of 32,400 images that represent six different hand gestures performed by four different individuals.

Preprocessing: Sadi et al. [26] proposed a method using YCrCb skin segmentation, Haar Cascade object detection, KCF tracker, and binary image conversion. Sahoo et al. [27] proposed a technique of thresholding based on depth (10 cm) with maximum area filtering, minimum–maximum normalization (Equation (3)), and a jet color map conversion. Zhou and Chen [28] have proposed a method for encoder-decoder segmentation using deep residual networks (DRN), atrous spatial pyramid pooling (ASPP), dilated spatial convolution (DSC), batch normalization, and data augmentation. Dang et al. [29] used HRNet pose estimation with keypoint normalization (per their formulation), and bounding box expansion. Padhi and Das [30] used MediaPipe detection with a BGR to RGB conversion. Mohamed et al. [31] used thresholding, region filling, and random vertical flipping. Wu et al. [32] used mosaic data enhancement and HSV gamut transformation. Bhavarthi et al. [59] used MediaPipe landmarks extraction with distance calculations (Euclidean distance, Equation (3)). Madaan et al. [60] used resizing to 224 × 224 with pixel normalization, rotation, zoom, and flipping. Awaluddin et al. [61] used on-the-fly augmentation with 30 times duplication, background replacement, geometric, brightness, temperature, and blur adjustment. Kumar

et al. [34] used preprocessing with horizontal flip, grayscale conversion, Gaussian blur, and thresholding.

Approaches: Gadekallu et al. [24] indicated that hyper-parameter tuning is a highly under-exploited area of research and would be extremely useful for real-time video processing. Bhushan et al. [25] reported that algorithmic performance can vary depending on the specific application, necessitating task-specific algorithm selection. Sadi et al. [26] identified difficulties caused by extreme lighting and significant variation in skin tone. Sahoo et al. [27] found that limited ability to extract keypoints from out-of-plane rotations limits LOO-CV performance of face-detection systems. Zhou and Chen [28] highlighted the need for lightweight models, while Dang et al. [29] noted that dark images or occlusions negatively impact keypoint extraction. Padhi and Das [30] suggested early stopping to improve training outcomes. Wu et al. [32] proposed optimization using Global-Sparse Convolution (GSConv) and pruning to reduce parameters. Jafari and Basu [33] emphasized that performance depends on the subset of features selected. Bhavarthi et al. [59] highlighted the importance of rigorous testing. Awaluddin et al. [61] reported that system effectiveness depends on the characteristics of training/testing datasets. Kumar et al. [34] noted confusion between similar gestures such as "S" and "R," and Tran and Nguyen [35] observed that user interface design and model stability can limit performance.

Challenges and Future Work: Key issues include: improving hyperparameter tuning [24], selecting task-specific algorithms [25], addressing extreme lighting and skin tone variation [26], mitigating out-of-plane rotation effects [27], developing lightweight models [28], handling occlusions [29], applying early stopping [30], pruning and parameter optimization [32], selecting optimal features [33], ensuring rigorous testing [59], managing dataset biases [61], minimizing gesture confusion [34], and improving user interface and model stability [35].

Machine Learning Techniques

Machine learning approaches combine traditional computer vision with classifiers, offering interpretable models and efficient computation.

Dataset: Nivash et al. [36] utilized 24,000 images across 20 gesture types from Kaggle, with 900 training and 300 testing images per gesture. Khaksar et al. [37] created a custom dataset of eight gestures from American, Indian, and Russian sign languages, with 10 images per gesture converted to 3D skeletal models using MediaPipe Hands.

Preprocessing: Nivash et al. [36] employed HSV color space analysis (Hue 315, Saturation 94, Value 37) for skin segmentation, background subtraction, binary image conversion, and distance transform for palm center localization. Khaksar et al. [37] used MediaPipe Hands for 21-landmark detection with 2.5D coordinates, computing 3D vector angles using dot product and a 2D method ignoring depth, validated against goniometer measurements.

Approaches: Nivash et al. [36] proposed a multi-layer system combining GestureNet (HSV-based segmentation) with Mobile FaceNet (triplet loss) for security, achieving YOLO Face 98.9% accuracy and an ultralight detector at 97.2% with 8 MB model files. Khaksar et al. [37] conducted evaluation of four classifiers on MediaPipe-extracted landmarks: Z-axis instability caused average accuracy of 86.7% (dropping to 41.5%). Bounds-based classifier achieved 96.25% with 81.4 ms latency, outperforming SVM (81.3%, 88 ms), ANN (77.5%, 196 ms), and linear regression (70%, 88 ms).

Challenges and Future Work: MediaPipe uses a normalized depth coordinate system which is unstable along the z-axis, limiting rotational robustness [37]. Environmental factors such as lighting, background, and skin tone can also affect gesture recognition accuracy [36]. Latency affects real-time performance. Future research should focus on: integrating voice commands, ensemble methods, emergency alert systems, fall detection [36],

comparing gesture identification algorithms quantitatively, applying clinical diagnostics, and optimizing for embedded deployment [37].

Rule-Based Techniques

Rule-based systems provide lightweight, interpretable solutions that map hand or facial landmarks to predefined commands based on logical thresholds and geometric conditions.

Dataset: Rule-based methods do not require extensive training data and are suitable for low-cost, low-latency embedded implementations. Huda et al. [38] developed an RGB camera dataset with 20 samples per gesture, achieving 99.17% accuracy. Extended work [39] expanded this to 100 samples per gesture (700 total) across varying hand sizes, achieving 98.14% accuracy with F1-scores greater than 0.9. Dragoi et al. [40] conducted user studies with 31 participants navigating a six-page PDF using five gestures, achieving 86.7% accuracy, 872 ms latency, and 4.03/5 ease-of-use. Ritu et al. [41] developed a facial gesture dataset with 100 trials per command under varying lighting and distance conditions, achieving mouth detection rates of 80–98% with motor response times of 0.5–1.2 s.

Preprocessing: Huda et al. [38,39] used MediaPipe Hands for 21-landmark detection, converting BGR to RGB and normalizing coordinates using image dimensions. A distance matrix was constructed using the Euclidean distance (Equation (3)), normalized by hand width (distance between keypoints 5 and 17). Thresholds (Thld1, Thld2, Thld3) were determined from 100 live hand samples capturing maximum, minimum, and mean values. Dragoi et al. [40] applied MediaPipe with OpenCV for real-time detection and implemented a cooldown mechanism to prevent repeated detection. Ritu et al. [41] utilized dlib and OpenCV for 68 facial landmarks, calculating the Mouth Aspect Ratio (MAR) and defining a central reference zone (60×50 pixels) for directional computation.

Approaches: Huda et al. [38] developed a distance-based mathematical model using thresholds Thld1 (maximum fingertip distance), Thld2 (minimum fingertip distance), and Thld3 (inter-finger distances). Logical conditions include: Stop (all $d(0, 8/12/16/20) >$ Thld1), Horn (all $<$ Thld2), and Drive ($d(0, 8) >$ Thld1 while others $<$ Thld2). Extended work [39] replaced fixed thresholds with dynamic ranges from 100 samples and introduced tolerance thresholds for directional movements. Dragoi et al. [40] used a geometric approach: right/left (direction of pointing finger), two fingers (index/middle extension), OK (thumb/index circle), and palm (all fingers extended), integrated with a local LLM (Ollama) for content generation. Ritu et al. [41] created a face-controlled system where MAR > 0.6 triggers mouth opening/closing; nose position relative to center zone determines movement direction; largest face in frame is prioritized.

Challenges and Future Work: Huda et al. [38,39] identified that fixed thresholds limit flexibility; dynamic ranges partially resolve this, but issues remain with lighting and background variations. Dragoi et al. [40] reported reduced precision ($\sim$90%) for left/right gestures due to small angle differences and variable response times (492–1554 ms). Ritu et al. [41] noted missing features: obstacle detection, waterproofing, wireless control, and requirement of $\sim 30^{\circ}$ head movement. Future work should address gesture speed detection, fall detection, emergency braking, online monitoring [38,39], expanded gesture vocabularies, adaptive recognition models, and local AI optimization [40]. Incorporating ultrasonic/LIDAR sensors, wireless control, and support for limited facial mobility would further improve usability [41].

2.2.2. Dynamic Vision

Dynamic gesture recognition identifies hand movements over time, considering trajectory, velocity, and temporal evolution. These systems process video sequences to extract spatio-temporal features, capturing both spatial configurations and temporal dynamics.

Deep Learning Techniques

Deep learning methods leverage spatio-temporal features from video sequences using 3D CNNs, recurrent networks, transformers, or hybrid models.

Dataset: Gonzalez Leon et al. [45] created a customized dataset consisting of 762 sequences (40 frames per sequence) of 6 different gestures using the Intel RealSense D435 (Intel Corporation, Santa Clara, CA, USA). Peral et al. [46] used the IPN Hand to obtain 4218 instances from 200 video clips (50 subjects). Riaz et al. [47] used a subset of 20BN-Jester that includes 30,000 videos (15 classes). Nguyen et al. [48] used IPN Hand to collect data for continuous gesture recognition with 13 gestures plus no-gesture (over 4000 samples). Miah et al. [49], and Mohammed et al. [50] have employed several skeleton-based datasets including: MSRA (17 gestures), DHG-14/28 (14/28 gestures), SHREC'17 (14/28 gestures), and LMDHG (13 gestures). Cui et al. [55] recorded a multimodal dataset with 10 users (each user performed 50 repetitions of a gesture). Chamalsha et al. [56] created an elderly-focused dataset that consisted of 560 samples (five gestures). Narayanan et al. [51] collected over 2000+ radar point clouds as snapshots of the hand (13 hand shapes). Wang et al. [52] obtained 7000 images of RTM/DTM/ATM through FMCW radar. Bremer et al. [53] recorded VR-based datasets that include both eye-tracking and EEG from 20 participants. Meghna et al. [58] designed a simulation environment for testing of navigation.

Preprocessing: The authors of Peral et al. [46] used four different methods for determining key frames: raw ($42 \times M$), Distances (Euclidean), DistAndTime (adding a self-distance), and DistTime (cross-frame distances), each defined in their work. Of these methods, they found that DistTime, with $M = 15$, produced the most accurate results. In addition to implementing affine transformations, contrast normalization, and adding Gaussian noise, Riaz et al. [47] improved the accuracy in training-validation by 25–45%, resulting in an accuracy of 79–99%. Nguyen et al. [48] took advantage of the MediaPipe skeleton and performed a linear interpolation on this skeleton to determine a TD-Net feature set (using joint-collection distances, normalized joint coordinates, and slow/fast-motion features). Miah et al. [49] subtracted the palm position of the first frame, then applied a position-embedding method using both sin/cos functions. Mohammed et al. [50] normalized their sequences temporally and spatially through scaling (0.8–1.2×), shift ($\pm$0.1), temporal interpolation, and random noise. Narayanan et al. [51] took radar data and created 5D point clouds. Bremer et al. [53] divided their data into 100 ms segments, then aligned all the data with a yaw transformation to provide a common reference frame, and finally analyzed EEG data through the use of a multitaper spectral density.

Approaches: González Leon et al. [45] have designed a lightweight 3D CNN model with three Conv3D layers that achieve 99.48% (RGB) and 99.18% (depth) with small loss of accuracy by using 1/2 or 1/4 of the frame rate. Peral et al. [46] have proposed the use of EUREKA in combination with MediaPipe and densely connected networks, reaching 87.5% on the IPN Hand with an average batch processing speed of 10 FPS. Riaz et al. [47] have proposed the use of a six-layer 3D-CNN in combination with an LSTM (512 units) to reach 97.5% validation accuracy and 97% test accuracy, using only 3.7M parameters. Nguyen et al. [48] proposed a three-model TD-Net pipeline for continuous recognition, reaching 84.98% of isolated accuracy and 40.10% Levenshtein accuracy with an inference time of 0.1 ms. Miah et al. [49] have proposed a multi-branch architecture that uses spatial–temporal, temporal–spatial, and general features, reaching 94.12% (MSRA), 92.00% (DHG), and 97.01% (SHREC'17). Mohammed et al. [50] have developed MMEGRN that combines four sub-networks (M3DCNN, MTCN, MConvLSTM, MMFN) with a weighted ensemble optimization technique, reaching 96.43% (SHREC'17), 92.18% (DHG), and 95.88% (LMDHG). Cui et al. [55] have developed a system that integrates gestures (98.2%), voice

(98.4%), and head posture (95.6%) with 3D head pose to 2D mapping, resulting in a <10 cm navigation error. Narayanan et al. [51] have introduced 3D deformable DETR for radar point clouds with 3D-GIoU loss, achieving 60.89% mAP@50 (20% improvement). Wang et al. [52] have combined F-RCNN with 3-DCGAN and AMSoftmax, achieving 80.8% mAP and 96.3% RQS. Bremer et al. [53] have evaluated and compared the performance of Transformers (78%) and LSTMs (76%) for detecting gaze intentions, as well as solving the Midas Touch problem with AR feedback.

Challenges and Future Work: Gonzalez Leon et al. [45] noticed that similar actions to zoom in/out can be confusing. Peral et al. [46] found that the lighting of the room and the angle at which a person's hands are oriented will influence the accuracy of the MediaPipe. Riaz et al. [47] was able to obtain good results for their model however they did note that there is a cost of resources required to process their model. Nguyen et al. [48] showed that the continuous recognition accuracy for their model was very low (40.1%). Miah et al. [49] pointed out that the rigid structure of graphs may limit how well the system can generalize across different data sets. Mohammed et al. [50] indicated that the number of nodes in an ensemble affects the performance of the system. Cui et al. [55] demonstrated that one of the major sources of complexity in the system is the wiring of the sensors. Chamalsha et al. [56] demonstrated that the performance of the system in the outdoors was significantly less than it was indoors (90.4% vs 92.7%). Narayanan et al. [51] were able to achieve a mean Average Precision of 60% with their model; however, they also noted that their model had a significant amount of confusion when recognizing gestures that are similar. Wang et al. [52] demonstrated that another challenge for their model was dealing with the large amount of intra-class variation present in many gestures. Bremer et al. [53] demonstrated that the behavior of people using the system would differ in a simulated environment as opposed to a real-world environment. Meghna et al. [58] demonstrated that their model could detect obstacles in a simulated environment with a 100% success rate; however, they also demonstrated that there were constraints on the usage of their model due to battery life. The next steps for this research area will include developing systems that can handle multiple persons [45]; developing adaptive recognition models [46]; optimizing the deployment of edge processing for the systems [47]; developing end-to-end systems [48]; developing adaptive graph learning models [49]; applying knowledge distillation techniques [50]; developing modular designs and conducting outdoor testing [55]; developing robust methods to optimize the systems for use in the outdoors [56]; developing systems that can learn continually and utilize larger datasets [51]; developing systems that have been trained on a wider variety of data [52]; conducting studies to determine how users adapt to new systems [53]; and deploying systems on hardware with a wide variety of user types [58].

Machine Learning Techniques

Machine learning techniques provide interpretable models that allow users to manage their own variability by designing features and validation mechanisms, which are particularly important for assistive technologies because hand tremors can significantly degrade the accuracy of gesture recognition.

Dataset: A human study was performed by Bandara et al. [54] with 25 wheelchair users aged 55–75 who were asked to perform twelve pre-defined gestures (five static and seven dynamic) using a Leap Motion sensor to capture the position of the fingertip and the palm. The data collection consisted of two experiments. In experiment one twenty participants (mean age of 63) were given an initial model without any validation mechanism. In experiment two another twenty participants (mean age of 59) were also provided a validation system.

Preprocessing: The working area of the Leap Motion sensor was defined as 30 cubic centimeters (cm^3) and divided into 216 rectangular areas (six by six by six) of five cubic centimeters (cm^3) per side. Each gesture generated a frequency map of hand features for each of the 216 rectangular areas. Three 2D activity maps were created by projecting these frequency maps onto the x-y plane, x-z plane, and y-z plane.

Approaches: A two-stage machine learning system was developed. Stage one is a baseline neural network classifier that has been shown to have 90 percent accuracy. This is followed by a second stage that utilizes a gesture validation mechanism. This gesture validation mechanism consists of a convolutional neural network (CNN) that has one convolutional layer (with 64 filters) followed by a max pooling layer, a dense layer (with 1024 neurons that utilize the rectified linear unit activation function), and a hidden layer (with 128 neurons). The input to this CNN is three 4×4 projected activity images. If the validation mechanism produces a different classification than the baseline mechanism, the system will automatically redirect to the second stage for additional classification attempts. The analysis of this two-stage system indicated that static gestures were tightly clustered regardless of the amount of hand tremors exhibited by each participant. Dynamic gestures however were very widely scattered. The Kappa coefficient improved after the addition of the validation mechanism. For static gestures the Kappa coefficient improved from 0.8994 to 0.9849, while for dynamic gestures the Kappa coefficient improved from 0.8479 to 0.984.

Challenges and Future Work: Person-to-person variability due to hand tremors causes significant variability in the frequency of occurrence of each feature within each rectangular area. Therefore, it is difficult to establish fixed feature ranges that will be effective across all users. Differences in tremor severity among individuals can result in static gestures exhibiting greater dispersion and, therefore, being similar to dynamic gestures. Therefore, future research efforts need to focus on developing an adaptive method of defining the size of each rectangular area based upon the specific characteristics of each individual's tremors, utilizing real-time compensations for tremors, expanding the population studied and including both larger and more diverse populations, and adapting the method of dividing the rectangular areas based upon the unique spatial requirements of each type of gesture.

2.2.3. Hybrid Vision

Hybrid vision-based control systems use combinations of computer vision, deep learning, and intelligent tracking to produce adaptive and context-dependent wheelchair navigation. These hybrid systems combine gesture recognition, object detection, lane detection, and tracking mechanisms.

Deep Learning Techniques

Hybrid systems combine multiple deep learning architectures to take advantage of the complementary strengths of these architectures in capturing spatial and temporal features. Specifically, CNNs can be combined with Recurrent Networks (RNNs) such as LSTM and GRU, or Attention Mechanisms.

Dataset: Camera-based gesture sensing was used by D'Souza et al. [42]. Gesture images and obstacle samples were used by Mahdin et al. [18] to develop object detection capabilities using the MS-COCO dataset. A multimodal visual dataset was developed by Abiraj et al. [43] that contained images of lanes, facial landmarks, gaze direction, and gesture sequences. A dynamic gesture video dataset was developed by He et al. [44] for gesture detection and tracking purposes using RGB cameras.

Preprocessing: D'Souza et al. [42] performed gesture segmentation isolating hand motion patterns. Mahdin et al. [18] applied image resizing and normalization. Abiraj et al. [43] used edge detection, color thresholding, and region-of-interest extraction. He et al. [44] employed Gaussian mixture modeling for skin detection, particle swarm optimization for feature selection, and adaptive data augmentation (brightness enhancement, geometric transformation).

Approaches: D'Souza et al. [42] implemented a camera-based hybrid pipeline in which a convolutional neural network classifies hand gestures and the recognized command is combined with on-board obstacle sensing to drive the wheelchair, providing an intuitive contactless interface. Mahdin et al. [18] coupled gesture-based directional control with a deep learning object detection module (trained on MS COCO) and GPS tracking, so that recognized gestures are overridden by collision avoidance logic when an obstacle is detected. Abiraj et al. [43] fused multiple visual cues, including hand gestures, facial landmarks, gaze direction, and lane detection, within a single multimodal controller, allowing the system to switch between gesture-driven and lane following navigation according to context. He et al. [44] combined an improved particle swarm optimization feature selection stage with a kernelized correlation filter (KCF) tracker to maintain robust hand tracking across frames, improving gesture tracking stability under motion. Collectively, these hybrid vision systems prioritize redundancy: gesture recognition supplies the primary command, while object detection, tracking, or lane following provide complementary safety and context awareness.

Challenges and Future Directions: Challenges include illumination variation, occlusion, and computational complexity. Gesture accuracy degrades under complex backgrounds; object detection models require optimization for embedded systems; tracking algorithms experience drift during rapid motion. Future research should focus on lightweight deep learning architectures, illumination-invariant vision models, adaptive multimodal fusion, and self-learning intelligent control systems.

3. Signal-Based Control

3.1. Mechanical Signal-Based Control Methodology

Signal-based control methodologies represent another major category of smart wheelchair control systems. In these approaches, wheelchair operation is based on the direct measurement of user intentions, including hand movements, eye movements, or other communication signals. Researchers have explored various mechanical signal-based interfaces to improve wheelchair navigation and user interaction. For instance, Patankar et al. [62] developed an Internet-of-Things-based wheelchair in which an ADXL335 accelerometer (Analog Devices, Inc., Wilmington, MA, USA) worn on the user's hand transduces hand tilt into directional commands, while Kalantri and Chitre [63] proposed an automatic wheelchair that maps accelerometer-derived hand-motion signals to forward, left, right, and stop commands. These mechanical-signal interfaces are attractive for low-resource settings because they require neither cameras nor model training, although their reliance on fixed tilt thresholds limits adaptability across users.

Beyond simple tilt mapping, more recent work decodes richer mechanical signals using machine learning. A representative example is the composite fiber aerogel pressure-sensing framework of Yang et al. [64], which is examined in detail below.

Dataset: The mechanical signal-based wheelchair control framework developed by Yang et al. [64] uses a composite fiber aerogel pressure sensor that converts the mechanical deformation of the backrest of the wheelchair into an electrical resistance variation. As the backrest of the wheelchair is compressed or released due to the movement of the user's back and shoulders, the resistance of the sensor varies accordingly. This variation in

electrical resistance is directly proportional to the magnitude of the mechanical deformation experienced by the sensor. As a result, the sensor provides a continuous time-series pressure signal that encodes information about the user's gestures and activities. Data is collected across multiple repetitions of the same gestures and activities to capture variability in the applied force and deformation behavior. For example, Yang et al. [64] demonstrated how their mechanical signal-based wheelchair control framework could be used to classify gestures and activities by training a CNN on a dataset of 2000 samples of pressure signals recorded over 5 days. Each sample represented a single gesture/activity execution, and the CNN learned to classify these samples based on their unique pressure signal characteristics. The results showed that the CNN achieved a classification accuracy of 98%, demonstrating its ability to effectively recognize a wide variety of user gestures and activities based on the pressure signal data produced by the sensor.

Preprocessing: Preprocessing plays a critical role in preparing the pressure signal data for use in the CNN. First, the raw electrical signals produced by the pressure sensor must be stabilized to remove transient fluctuations and electrical noise that arise during rapid mechanical loading and unloading. Second, the sensor exhibits a linear relationship between applied pressure and measured electrical resistance within a well-defined operational range. Calibration is performed to ensure that this relationship remains consistent across all recordings, thus, preserving proportionality required for effective learning. Next, the continuous pressure signals are segmented into discrete temporal windows corresponding to complete gesture cycles. The segmentation of the signals allows the model to capture dynamic deformation behavior, and to learn to distinguish between different gestures and activities. Finally, the segmented and normalized signals are converted to structured tensor representations that are suitable for processing by convolutional neural networks, transforming raw mechanical deformation into hierarchical learning-ready input [64].

Approach: The system employs a CNN to automatically extract discriminative spatial–temporal features from the processed pressure signal tensors. The workflow begins with feeding normalized time-series data into convolutional layers, where learnable filters detect localized deformation signatures associated with different gestures. Nonlinear activation functions enhance feature separability, while pooling operations reduce dimensional redundancy and improve robustness to minor signal variations. The resulting feature maps are then passed through fully connected layers that perform high-level classification, and a softmax layer produces probability distributions corresponding to gesture categories. Through this hierarchical feature extraction process, the model eliminates the need for manual feature engineering and achieves gesture recognition accuracy approaching 98% [64].

Challenges and Future Directions: Although the system's accuracy is high as well as sensitive, it faces a variety of challenges. Some of these include cross-talk between multiple fingers that move at the same time when inputting the system, fatigue over time of materials that compress repeatedly, and the computational demands that come from implementing convolutional neural networks (CNNs) on embedded hardware. User applied forces can vary, causing potential distribution shift effects in the generalization of the model. Therefore, future work will be focused on developing light-weight Deep Learning Architectures that allow for real-time embedded deployments, adaptive recalibration techniques to account for sensor degradation, and multi-modal integration techniques that combine mechanical sensing, motion, and/or biological sensing to enhance robustness and context-awareness.

3.2. Rule-Based Techniques

Rule-based mechanical control systems are based upon a deterministic relationship between the physical motion of the user and the wheelchair commands that are issued, but they do not require model training to provide this mapping. In most cases, rule-based

methods rely on accelerometer derived tilt signals to determine the direction of user hand orientation and then translate that information into directional wheelchair movement. When compared to deep-learning methods, rule-based methods are computationally inexpensive, fast, and can easily be implemented in an embedded fashion. However, the success of rule-based methods is highly dependent on the accurate calibration of thresholds and the reliable measurement of sensor data [62,63].

Dataset: The rule-based mechanical control systems utilized the accelerometer derived tilt measurements to determine the directional commands of the wheelchair [62,63]. As opposed to the adaptive deep-learning method, the dataset consisted of the analog voltage output of the ADXL335 accelerometer, representing hand orientation along the x, y, z axes.

Preprocessing: In addition to a focus on acquiring reliable, consistent signals, Rule-Based System (RBS) pre-processing also focuses less on extracting features, as opposed to machine learning-based feature extraction [62,63]. The analog output of the accelerometer is then converted to digital format via analog-to-digital conversion within the microprocessor. Aspects of the system that are of interest can be isolated via axis selection; directionally relevant aspects are most commonly selected from the x-axis and y-axis for 2D/planar motion control. To reduce jitter, signal smoothing is added to prevent motion detection from being disrupted by hand movement or vibration of the sensor. To establish a predefined angle boundary map, threshold calibration is performed and directional command will be issued when the tilts exceed this predefined angle boundary.

Approaches: The rule-based control system uses determinative logic that compares continuously digitized tilt values against predetermined calibrated limits by means of conditional statements contained within the microprocessor's firmware [62,63]. If a threshold condition has been met, the microprocessor enables relay circuits or motor drives, controlling both the speed and direction of the DC motors attached to the wheelchair. The above method allows for quick responses and as little computation as possible for use on an embedded platform.

Challenges and Future Directions: Although Rule-Based Systems (RBS) are relatively easy to design and operate computationally efficiently, RBSs have limitations in terms of their adaptability and tolerance for noise and variations in user input. The fixed thresholds used by RBSs to determine when a user is using a specific gesture or interface function may not be generalizable to users who exhibit differences in the way that they move. Therefore, future work will need to develop and evaluate techniques that will allow the thresholds to adapt based on the user's interaction patterns and/or use of learning-based classifier algorithms as well as better sensor calibration methods to improve the overall robustness of the system without sacrificing the required responsiveness to the user's interactions.

3.3. Biological Signal-Based Control Methodology

Surface electromyography (sEMG), a type of neuromuscular signal-based upon the electrical activity generated by muscle contraction and represented as cumulative motor unit action potentials, represents the basis for biological signal-based wheelchair control systems that convert these signals into navigation commands. However, because sEMG signals are nonlinear, nonstationary, and highly sensitive to both physiological and environmental disturbances, they cannot be directly modeled using traditional mathematical techniques; therefore, reliable mapping of muscle activation patterns to wheelchair motion commands can only be achieved through machine learning models. The overall methodological approach to achieving this mapping includes signal acquisition, processing, feature extraction, machine learning model prediction, and finally, command mapping, thereby enabling users to interact with the wheelchair in an intuitive manner utilizing their neuromuscular activity.

Machine Learning Techniques

The use of statistical learning algorithms for interpreting the muscle activation patterns (sEMG), and translating those patterns into commands for navigating a wheelchair is based upon machine-learning biological control systems. These types of biological control systems compare time varying sEMG signals to determine both an individual's gesture intention and their gesture execution intensity. Unlike the rule-based biological control systems previously developed, the use of machine-learning allows for greater adaptability, increased accuracy, and increased robustness against variation in the sEMG signals captured from the individual performing the gestures. However, the overall effectiveness of the biological signal-based wheelchair control systems described above are significantly dependent upon the data pre-processing quality and the representation of the features in the data.

Dataset: Both biological signal-based wheelchair control systems proposed by Iqbal et al. utilized datasets consisting of multi-channel surface electromyography (sEMG) signals collected from forearm muscles using a Myo armband equipped with eight dry electrodes. These signals represent the electrical activity generated during muscle contractions and are influenced by factors such as contraction intensity, gesture execution speed, and electrode placement. In one study [65], sEMG signals were recorded in real time while participants performed multiple hand gestures for wheelchair navigation at a sampling rate of 200 Hz. In another related study [66], participants repeatedly performed gesture sequences to collect sEMG data containing realistic variations in muscle activation caused by differences in individual physical abilities. Consequently, the collected time-series biosignals effectively captured neuromuscular behaviors, forming the foundation for machine learning-based gesture recognition and wheelchair control.

Preprocessing: Preprocessing is vital to signal-based biological systems due to noise, motion artifacts and baseline drift present in raw sEMG signals [65,66] and it starts with signal conditioning. Filtering of sEMG signals starts with a notch filter to eliminate low frequency interference from power lines and DC offsets, followed by high pass Butterworth filtering to eliminate low frequency movement artifacts and increase clarity of the sEMG signals. sEMG signals are not stationary so after filtering the sEMG signals are segmented into overlapping windows of time representing individual phases of muscle activation. Time domain descriptors are used for feature extraction such as root mean square (RMS), mean absolute value (MAV), etc., to measure the amplitude of the sEMG signals and intensity of the contraction during each window. After feature extraction, each descriptor is normalized to account for inter-subject variability and location of electrodes on the subject to improve the generalizability of the model. Overall through filtering, segmentation, feature extraction, and normalization, preprocessing transforms raw biosignals into structured sets of features for use in machine learning.

Approaches: The process for gesture recognition and control of the wheelchair utilizes supervised learning algorithms to translate the sEMG feature vectors into the specific wheelchair movements. In Iqbal et al. [65] they utilize K-Nearest Neighbors (KNNs) and an ensemble classifier to identify the gestures from the extracted sEMG feature vectors. Using KNNs, the algorithm recognizes the gesture class based upon the similarity among the neighboring vectors; and utilizing the ensemble classifier, it combines multiple decision trees to increase the robustness of the classifications. The KNN and ensemble classifiers begin the process of recognizing the gesture class from the sEMG feature vectors that have been segmented from the biosignals. Once the supervised training has utilized the labeled gesture data, the trained classifier will classify the gesture class in real time and then translate that gesture class into the directional controls of the wheelchair. In contrast, Iqbal et al. [66] use a regularized linear regression model to establish a direct relationship between the sEMG feature vectors and the output of the wheelchair, allowing for the proportional

and simultaneous control of the wheelchair as opposed to a discrete classification of the gesture. The regression model is able to minimize the prediction error using least squares optimization, while also providing the necessary regularization to allow for smoother and more natural movement. Both methodologies require the system to proceed from the acquisition of signals, through the extraction of features from those signals, to the inference of machine learning and the creation of commands for the wheelchair, thereby facilitating a real-time human–machine interface using neuromuscular signals.

Challenges and Future Directions: Biological signal-based control systems have challenges due to signal variation from electrode position, muscle fatigue, and physiological differences in individuals using them. The non-stationarity of sEMG signals complicates feature extraction and the ability for a model to generalize, and motion/noise artifact can degrade classification accuracy. Gesture classification provides either discrete control or no control for continuous movement, and regression-based approaches are computationally expensive and require accurate calibration. Therefore, future research will likely need to address these issues through user independent learning models; multimodal integration (with mechanical or vision-based sensing); adaptive calibration of the system; and/or lightweight, real-time machine learning implementations to improve both the usability and robustness of biological signal-based wheelchair control systems.

3.4. Visual and Optical Signal-Based Control Methodology

Human–machine interface using visual- and optical-signal-based wheelchair control systems can be made possible by the analysis of gesture patterns and/or optical signal variation to generate navigation commands from visual gestures and/or optical signal variations. In contrast to the mechanical and biological methods, which rely on camera-based video processing or light-modulated sensing for detecting gesture-induced changes in visual or optical signals, this method utilizes machine learning for the analysis of gesture features detected in visual frames or optical signal variations to allow real-time wheelchair operation by non-contact interfaces. The typical process flow is composed of signal capture by cameras or optical sensors, enhancement of signal quality, feature detection, machine-learning-based gesture recognition, and generation of commands for wheelchair navigation.

Machine Learning Techniques

The visual- and optical-control systems utilize machine learning to analyze hand gestures captured by means of visual sensing or optical sensing and transform the captured hand gestures into navigation commands. The primary advantages of machine learning are user comfort and increased accessibility due to the absence of physical contact between the user and the machine. The primary advantage of machine learning over rule-based visual methods is its ability to provide increased robustness in terms of noise, lighting variation, and gesture variability. However, the quality of the pre-processing and accuracy of the feature extraction are critical factors affecting the performance of the system.

Dataset: Nithya et al. [67] proposed the visual-gesture-based wheelchair control system. The dataset consisted of continuous video frames of a camera capturing the user's hand-gesture movements. The captured visual data contained information about the user's hand-gesture motion trajectories, shapes, and timing; similar to the optical-signal-based gesture-recognition system proposed by Liang et al. [68], the dataset consisted of VLC-sensed light-intensity variations, where user hand gestures modulated optical signals received at photodiode sensors. The optical-signal-based dataset had multiple gesture types, including directional motion, rotational motion, and positional changes. The gestural types were well balanced and allowed for machine-learning-based classification with an accuracy

of approximately 95.7%. Both the visual- and optical-signal-based datasets represented the spatial-temporal characteristics of gestures either through the captured image frames or optical signal variations and, thus, provided a basis for machine-learning-based visual and optical gesture recognition.

Preprocessing: Preprocessing in vision and optics-based gesture systems primarily includes improving the quality of gesture signal and developing important gestures from visual or optic data. Image processing is applied to the frames that are taken from the camera in the camera-based gesture system of Nithya et al. [67]. This allows gesture trajectories and gesture patterns to be isolated from the background of each frame. The frames are processed sequentially to allow the hand motion to be detected. Recognized gesture patterns are sent to the control unit for command execution. On the other hand, Liang et al. [68] utilize a structured signal processing approach in their VLC-based optical gesture system. The raw optical signals are filtered using a low-pass Butterworth filter to remove ambient light noise prior to being normalized to ensure the same level of sensitivity exists across all sensing channels. The data dimension is reduced through downsampling while maintaining gesture characteristics. Sliding window segmentation is employed to separate gesture specific optic patterns for further analysis. Gesture features are extracted to include both temporal and spatial gesture characteristics (i.e., amplitude modulation, etc.) allowing for robust classification via machine learning techniques. In this way, the preprocessing steps (image filtering, signal conditioning, segmentation, feature extraction) convert the raw visual and optic data into machine learning-based gesture recognition compatible structure.

Challenges and Future Directions: Visual and optical gesture recognition systems face challenges related to lighting variability, background complexity, and occlusion in camera-based systems, as well as sensitivity to optical interference and environmental light variations in VLC-based systems. Additionally, real-time processing requirements and computational complexity may limit embedded deployment. Future research should focus on robust low-light vision algorithms, adaptive optical signal calibration, multimodal fusion combining visual, mechanical, and biological sensing, and lightweight real-time machine learning architectures to enhance reliability and usability of visual and optical signal-based wheelchair control systems.

3.5. Audio and Multi-Modal Signal-Based Control Methodology (Hybrid)

Audio and multi-modal, signal-based wheelchair control systems are developed to provide users with an intelligent control system that provides the user with a hands-free approach to controlling their wheelchair through combining voice recognition and intelligent monitoring using environmental sensors. These types of control systems have the advantage of being able to combine multiple methods of control for example speech interaction, sensor-based perception, and artificial intelligence to improve safe navigation, safety, and overall user independence compared to single-modality control systems. Pasam et al. [69] propose a control system that integrates IVRS technology, AI driven voice recognition, IoMT physiological monitoring technology, and autonomous navigation technology into one system creating a hybrid control framework that can provide simultaneous real-time interaction, environmental awareness and health care monitoring. The methodology of the system proposed is composed of the following steps: Voice and sensor data collection and preprocessing, Multimodal signal interpretation, Intelligent decision making, and Command Execution.

Hybrid Techniques

Multimodal hybrid audio-control systems, which utilize artificial intelligence (AI) to integrate a combination of speech input, physiological monitoring, and environmental sensing can provide intelligent wheelchair navigation systems that allow users to operate their wheelchairs hands free while also enhancing safety with obstacle detection and/or monitoring the user's health in real time. Additionally, AI multimodal systems are more robust than single-modality voice-controlled systems due to increased awareness of context. However, the overall functionality of these systems is dependent upon high accuracy of speech recognition, reliability of sensor data from the sensors being used, and sufficient processing speed to perform tasks in real time.

Dataset: A hybrid audio and multimodal control system has been presented by Pasam et al. [69] which utilizes a dataset that includes voice commands, physiological sensor signal data and environmental sensor data, all of which have been collected during use of a wheelchair. Physiological sensor signals captured using IoMT (Internet of Medical Things) include various physiological parameters including temperature, heart rate, ECG and SpO_2; Environmental sensing data was obtained from ultrasonic sensors. All of these multimodal data sources provide the same three things: Combined User Intent; Health Status of the user; and Context of Environment, thereby allowing the autonomous wheelchair to make decisions based on both user intent and environment and monitor its own status.

Preprocessing: The pre-processing stage of the proposed hybrid architecture provides an appropriate platform for processing the audio and sensor signals, to support reliable interpretation and make decisions. Pre-processing involves applying noise reduction on the audio signals by the microphone to eliminate ambient interference and to enhance the accuracy of speech recognition in various environments. After eliminating noise from the audio signals, the filtered audio signals are subjected to voice recognition by employing AI-based algorithms to interpret spoken commands as textual inputs to provide semantic understanding of the user's intentions. Physiological signals that have been collected by the IoMT sensors are conditioned and normalized so that readings such as temperature, heart rate, ECG and SpO_2 are consistent across all readings. Environmental sensing data is calibrated by the ultrasonic sensor(s) to detect possible obstructions and to assist with accident prevention while navigating. Audio, physiological, and environmental data will be combined to create a single multimodal input allowing for coordinated decision making. Through the process of noise filtering, feature extraction, normalization, and multimodal synchronization, preprocessing allows for proper interpretation of user commands and environmental conditions for the safe and intelligent operation of the wheelchair.

Approaches: The Hybrid Control Architecture combines speech recognition, autonomous navigation, and multimodal sensor fusion to provide Intelligent Wheelchair Operation (IWO) [69]. Spoken commands captured via a microphone convert to digital signals that are then analyzed using AI-based speech recognition to identify user intent. The decoded command is sent to the central controller (a Raspberry Pi; Raspberry Pi Ltd, Cambridge, UK), where multimodal data (voice, physiological sensors, environmental sensors) is integrated. The system evaluates obstacle distance to prevent collisions and simultaneously monitors patient health parameters. Once both environmental and physiological conditions are deemed safe, the central controller generates motor control signals that will drive the wheelchair along predefined navigation paths. Through continuous sensor feedback, the system can dynamically adjust its motion to form a closed-loop control system.

Challenges and Future Directions: In addition to the issues associated with multimodal input described above, multimodal and audio-based control systems also encounter several problems as well: environmentally related noise reduces the accuracy of speech recognition

by a control system; heterogeneous sensor data is difficult to integrate due to differences in timing and/or calibration; and integrating multimodal (multi-sensory) information into real-time processing is computationally intensive. Multimodal input has a bright future ahead of it, with many possibilities including real-time cloud-based monitoring of multimodal user interactions, voice interaction in multiple languages, adaptive multimodal fusion, lightweight AI models optimized for embedded applications, predictive obstacle detection, and intelligent health monitoring (for additional enhancements). A comparative summary of the signal-based wheelchair control systems reviewed above is presented in Table 4.

Table 4. Comparative summary of signal-based wheelchair control systems (mechanical rows with non-verifiable citations removed; author names corrected).

Author	Category	Sensor/Modality	Dataset	Preprocessing	Approach	Performance
Yang et al. [64]	Mechanical	Fiber aerogel pressure sensor	2000 pressure samples (5 days)	Stabilization, segmentation, normalization	CNN	98% accuracy
Patankar et al. [62]	Rule-Based	ADXL335 Accelerometer	Tilt voltage signal	ADC, threshold calibration	Deterministic logic	Directional movement
Kalantri and Chitre [63]	Rule-Based	Accelerometer + Wireless	Hand motion signal	Encoding, threshold mapping	Firmware-based logic	Forward/Left/ Right/Stop
Iqbal et al. [65]	Biological	Myo Armband (sEMG)	Real-time gesture dataset	Filtering, segmentation, RMS/MAV	KNNs + Ensemble	High gesture recognition
Iqbal et al. [66]	Biological	Myo Armband (sEMG)	Multi-user time-series dataset	Butterworth filter, normalization	Regularized regression	Smooth proportional control
Nithya et al. [67]	Visual	Camera-based sensing	Gesture motion frames	Frame processing, gesture extraction	ML-based classification	Gesture detection
Liang et al. [68]	Optical (VLC)	Photodiode + VLC	Balanced multi-gesture dataset	Filtering, segmentation, normalization	ML classification	95.7% accuracy
Pasam et al. [69]	Hybrid	Voice + IoMT + Ultrasonic	Multimodal dataset	Noise filtering, normalization, fusion	AI-based multimodal fusion	Voice navigation + monitoring

4. Discussion

4.1. Cross-Cutting Synthesis Across Modalities

Sections 2 and 3 described each modality in turn. To avoid a purely study by study narrative and to enable comparison across modalities, this subsection re-integrates the evidence along the dimensions that matter most for deployment: sensing principle and placement, recognition strategy, validation maturity, target population, and risk of bias. Table 5 summarizes the modality families against these dimensions, and Table 6 provides the underlying study-level coding.

A consistent pattern emerges from this cross-cutting view. Reported accuracy is inversely related to validation realism: the highest figures (>99%) come almost exclusively from offline evaluation of static gestures on public image datasets with healthy subjects, whereas studies that tested real prototypes, real users, or the target clinical population report lower but more trustworthy performance. Sensor- and signal-based modalities trade the lighting/occlusion fragility of vision for a calibration and donning burden, while hybrid systems consistently use one modality as the primary command channel and the others for safety and context. Because performance figures are not directly comparable across studies (they differ in task, number of classes, subjects, dataset, and protocol), the tables report these confounders alongside each accuracy value rather than ranking studies by headline accuracy.

Table 5. Cross-modality synthesis of gesture- and signal-based wheelchair control.

Modality Family	Typical Sensing & Placement	Key Strengths	Key Limitations	Highest Validation Maturity Observed
Vision—static (hand pose)	RGB/webcam; external or chair-mounted	Contactless; no wearable; high in-lab accuracy; MediaPipe enables real-time landmarks	Illumination, background, skin-tone and occlusion sensitivity; z-axis instability; mostly offline	Real-world user study (Drăgoi [40]); elderly testing (Sadi [26])
Vision—dynamic (trajectory)	RGB/depth/radar/EEG; external	Captures motion and temporal commands; richer command set	High compute; low continuous-recognition accuracy; simulation-to-reality gap	Indoor/outdoor prototype (Chamalsha [56])
Sensor—wearable motion/EMG	Myo, IMU, flex, capacitive; body-worn	Robust to lighting/occlusion; works when the hand is not visible; embeddable	Donning and per-user calibration; electrode/IMU shift; inter-subject variability	Real-time path test (Rusydi [15]); clinical population (Bao [9], Gopal [23])
Mechanical signal (pressure/tilt)	Aerogel pressure, ADXL335 tilt; chair- or body-mounted	Ultra-low-cost; rule variants need no training; easily embedded	Cross-talk; material fatigue; fixed thresholds limit adaptability	Lab prototype (Yang [64], Patankar [62])
Biological sEMG control	Myo sEMG; forearm body-worn	Serves severe upper-limb impairment; supports proportional control	Non-stationary signal; calibration burden; compute cost	Real-time prototype (Iqbal [65,66])
Optical/audio–multimodal	VLC photodiode, camera, voice + IoMT; external/onboard	Hands-free; integrates health and environment monitoring	Ambient-light/noise sensitivity; sensor-fusion timing and calibration	Lab prototype (Liang [68], Pasam [69])

To make the evidence base auditable, Table 6 codes every included study by sensing modality and placement (portable/body-worn, chair-mounted, or external), number of gesture classes, sample size, population (healthy, motor-impaired/clinical, elderly, or not reported), validation protocol and maturity (offline, simulation, lab prototype, or real-world), and an overall risk-of-bias (RoB) rating. The RoB rating uses four transparent signals: inclusion of the target population; evaluation beyond offline classification; use of between-subject or cross-validation rather than single-session resubstitution; and an adequately sized, clearly reported sample. Studies satisfying most signals are rated Low; healthy-subject, offline-only, single-dataset studies are rated Medium; and studies with very small or unreported samples, simulation-only evidence, or no reported validation are rated High.

Aggregating the coding in Table 6 yields the validation profile summarized in Table 7. Only 17 of 72 studies (23.6%) reported any real-world or field evaluation, 11 (15.3%) involved motor-impaired or elderly participants, and just 6 (8.3%) explicitly addressed pathological-tremor or unintended-movement compensation. The large majority of evidence is, therefore, offline classification on healthy subjects, which explains the gap between headline accuracy and demonstrated clinical readiness.

Table 6. Study-level characteristics, validation maturity, and risk-of-bias (RoB) assessment of the included studies. Population: H = healthy, I = motor-impaired/clinical, E = elderly, NR = not reported. Maturity tags: Off = offline, Sim = simulation, Pro = lab prototype, RW = real-world. RoB: L = low, M = medium, H = high. #Cls = number of gesture classes; #Subj = number of subjects.

Study	Sensor/Placement	#Cls	#Subj	Pop	Validation & Maturity	RoB
Sensor-based—deep/machine learning						
Vásconez [5]	Myo + G-Force; body-worn	11	85	H	Offline, user-specific train/val/test (Off)	M
Nogales & Benalcazar [6]	Leap Motion; body-worn	5	56	H	Offline, manual vs. auto features (Off)	M
Wang [7]	9-axis IMU; body-worn	6	20	H	Offline, EKF fusion (Off)	M
Zhao [8]	EMG + MediaPipe; body-worn	NR	NR	H	Offline architecture comparison (Off)	M
Bao [9]	sEMG; forearm body-worn	6	8	I	Offline; post-stroke patients (Off)	M
Zhang [10]	sEMG (Myo); body-worn	18/6	10/10	H	Offline, NinaPro DB5 + private (Off)	M
Fernandez [11]	Capacitive glove; body-worn	ASL	12 k smp	H	Offline (Off)	M
Kateb [12]	Textile capacitive; body-worn	12	NR	NR	Offline, *k*-NN (Off)	H
Kaur [13]	Gesture + fingerprint; mixed	NR	230 img	NR	Offline; biometrics (Off)	H
Zhang [14]	Myo sEMG; body-worn	5	6	H	Offline, cross-user transfer (Off)	M
Rusydi [15]	Flex + gyro; body-worn	5	20/6	H	Real-time 58.8 m path test (RW)	M
Singh [16]	MPU-6050; body-worn (edge MCU)	6	NR	H	On-device inference, lab (Pro)	M
Anam [17]	Myo sEMG; body-worn	5	5	H	Offline + real-time (errors rose) (Pro)	M
Gopal [23]	12-ch EMG; body-worn	10	4	I	Offline; transradial amputees (Off)	M
Sensor-based—rule-based (mechanical/inertial)						
Mahdin [18]	Gyro + ultrasonic + GPS; chair-mounted	dir.	50 trials	NR	Lab, lighting-varied trials (Pro)	M
Calado [19]	Glove + IMU; body-worn	10	5	H	Offline, Italian Sign Language (Off)	M
Chen [20]	Leap Motion; external	5 zones	20	H	Prototype + ease-of-use survey (Pro)	M
Islam [3]	MPU6050 + ultrasonic; body/chair	5	400 trials	NR	Prototype + YOLOv8 obstacle (Pro)	M
Rambabu [21]	MEMS acc.; body-worn	NR	NR	NR	Prototype, no metrics (Pro)	H
Balaji [22]	Head MPU6050; head-mounted	NR	NR	NR	Prototype, ESP-NOW (Pro)	H
Vision-based—static gestures						
Gadekallu [24]	RGB; external	10	20 k img	NR	Offline, Kaggle (Off)	M
Bhushan [25]	Sign-MNIST; external	24	24 k img	NR	Offline (Off)	M
Sadi [26]	Custom RGB; external	gestures	700 eld.	E	Tested on 700 elderly users (Pro)	M
Sahoo [27]	MU + HUST-ASL; external	36/34	5/10	H	Offline (Off)	M
Zhou & Chen [28]	OUHANDS; external	10	23	H	Offline, segmentation (Off)	M
Dang [29]	HANDS/SHAPE; external	15/32	dataset	NR	Offline (Off)	M
Padhi & Das [30]	HaGRID; external	18	1.9 k img	NR	Offline, MediaPipe (Off)	M
Mohamed [31]	Custom; external	44	105 k img	NR	Offline (Off)	M
Wu [32]	Custom; external	14	2.85 k img	NR	Offline, YOLOv5s (Off)	M
Jafari & Basu [33]	Multi-dataset; external	6+	large	NR	Offline, 6 datasets (Off)	M
Bhavarthi [59]	Custom (wheelchair); external	5	custom	NR	Lab prototype (Pro)	M
Madaan [60]	Kaggle subset; external	5	2.6 k img	NR	Offline (Off)	M
Awaluddin [61]	Green-screen; external	dataset	synthetic	NR	Offline, background-swap aug. (Off)	H
Kumar [34]	Custom; external	5	60 k img	NR	Offline, varied lighting (Off)	M
Tran & Nguyen [35]	Custom; external	6	4	H	Offline (Off)	M
Nivash [36]	RGB + face; external	20	24 k img	NR	Offline, face security (Off)	M
Khaksar [37]	MediaPipe (21 LM); external	8	80 img	H	Offline; goniometer-validated (Off)	M
Huda [38]	RGB; external	gestures	20/gest	H	Lab prototype, 99.17% (Pro)	M
Huda [39]	RGB; external	7	700 smp	H	Lab prototype, dynamic ranges (Pro)	M
Drăgoi [40]	Laptop cam; external	5	31	H	Real-world user study + LLM (RW)	L
Ritu [41]	Webcam (face); external	face cmds	100 trials	NR	Prototype, lighting/distance (Pro)	M
D'Souza [42]	RGB; external	gestures	NR	NR	Lab prototype (Pro)	H
Abiraj [43]	Face + gesture + lane; external	multi	NR	NR	Lab prototype, multimodal (Pro)	H
He [44]	RGB dynamic; external	gestures	video	NR	Offline, PSO + KCF tracking (Off)	M

Table 6. *Cont.*

Study	Sensor/Placement	#Cls	#Subj	Pop	Validation & Maturity	RoB
Vision-based—dynamic gestures						
González León [45]	RGB + depth; external	6	762 seq	H	Offline, 3D CNN (Off)	M
Peral [46]	IPN Hand; external	13	50	H	Offline (Off)	M
Riaz [47]	20BN-Jester; external	15	large	NR	Offline (Off)	M
Nguyen [48]	IPN Hand; external	13 + 1	4 k+	NR	Offline, continuous (Off)	M
Miah [49]	MSRA/DHG/SHREC; external	14–28	skeleton	NR	Offline (Off)	M
Mohammed [50]	SHREC/DHG/LMDHG; external	14–28	skeleton	NR	Offline, ensemble (Off)	M
Narayanan [51]	Radar; external	13	2 k+	NR	Offline, point cloud (Off)	M
Wang [52]	FMCW radar; external	gestures	7 k img	NR	Offline (Off)	M
Bremer [53]	VR + EEG; head-worn	gaze	20	H	VR simulation (Sim)	H
Bandara [54]	Leap Motion; external	12	25	E	Two experiments, elderly users (Pro)	L
Cui [55]	Gesture + voice + head; mixed	gestures	10	H	Prototype, <10 cm nav. error (Pro)	M
Chamalsha [56]	RGB webcam; external	5	560 smp	E	Indoor/outdoor prototype (RW)	M
Sirisati [57]	RGB video; external	Jester sub.	dataset	NR	Offline, YOLO + LSTM (Off)	M
Meghna [58]	Camera + LiDAR; onboard	sim	simulation	NR	Simulation only (Sim)	H
Signal-based—mechanical/biological/optical/multimodal						
Yang [64]	Fiber-aerogel pressure; chair-mounted	gestures	2000 smp	NR	Lab prototype, 5-day capture (Pro)	M
Patankar [62]	ADXL335 tilt; body-worn	4	NR	NR	Lab prototype (Pro)	M
Kalantri & Chitre [63]	Accelerometer; body-worn	4	NR	NR	Prototype (Pro)	H
Iqbal [65]	Myo sEMG; body-worn	gestures	NR	NR	Real-time KNNs/ensemble (Pro)	M
Iqbal [66]	Myo sEMG; body-worn	proportional	multi-user	NR	Real-time regression (Pro)	M
Nithya [67]	Camera; external	gestures	frames	NR	Prototype (Pro)	M
Liang [68]	Photodiode (VLC); external	multi	balanced set	NR	Offline, 95.7% (Off)	M
Pasam [69]	Voice + IoMT + ultrasonic; onboard	voice	NR	NR	Prototype, fusion (Pro)	M

Table 7. Validation profile of the 72 included studies (counts derived from Table 6).

Validation Dimension	Studies	Share
Evaluated beyond a controlled laboratory (prototype field use/real-world)	17/72	23.6%
Included motor-impaired or elderly participants	11/72	15.3%
Explicitly addressed pathological tremor/unintended movement	6/72	8.3%
Reported between-subject or cross-validation	29/72	40.3%
Offline classification only (no hardware or user deployment)	38/72	52.8%

This systematic review synthesized 72 studies published between 2022 and 2026 on gesture recognition systems for intelligent wheelchairs, encompassing vision-based, sensor-based, and hybrid approaches. The analysis reveals significant technological advances alongside persistent challenges that must be addressed to enable real-world deployment, particularly in low-resource settings like Bangladesh. Vision-based approaches have demonstrated remarkable progress, with deep learning architectures achieving classification accuracies exceeding 99% in controlled environments. MediaPipe-based systems offer lightweight landmark extraction suitable for real-time applications, while convolutional neural networks and transformer architectures provide robust spatial-temporal modeling. However, the review exposes a critical disconnect between laboratory performance and real-world reliability. Systems achieving 99% accuracy in controlled settings show significant degradation under variable lighting, complex backgrounds, and occlusions, representing a fundamental barrier to clinical adoption. Sensor-based methodologies employing EMG and IMU technologies offer advantages for users with limited hand visibility or motor control, with deep reinforcement learning and transfer learning approaches demonstrating promise for personalized adaptation. Yet sensor variability, electrode placement sensitivity, and the challenge of pathological tremor compensation remain unresolved. Hybrid multimodal frameworks combining vision, sensor, and audio inputs offer the most comprehensive

approach to robustness, achieving complementary strengths while mitigating individual modality weaknesses. The integration of gesture recognition with obstacle detection, health monitoring, and autonomous navigation reflects growing recognition that effective assistive technology must address the holistic needs of users rather than isolated control functions.

Limited real-world validation emerges as the most pervasive limitation across all studies. Of the 72 included studies, only 17 (23.6%) reported any evaluation beyond a controlled laboratory environment, and only 11 (15.3%) involved participants with diagnosed motor impairments; the per-study basis for these counts is given in the study-characteristics and risk-of-bias assessment (Table 6) and summarized in Table 7. This validation gap undermines generalizability claims and obscures failure modes that would manifest in daily use. Insufficient handling of pathological tremor constitutes a major clinical limitation, with only six studies explicitly addressing motor impairment compensation mechanisms despite neurological conditions causing mobility impairments frequently co-occurring with upper-limb movement disorders. Environmental robustness deficits persist despite advances in deep learning architectures, with illumination variation, background complexity, and humidity effects on sensors remaining inadequately addressed. Cultural and contextual misalignment represents a largely overlooked dimension, as gesture vocabularies derived from Western sign language conventions may not align with culturally intuitive gestures in South Asian contexts, and cost considerations, infrastructure limitations, and local manufacturing constraints are rarely incorporated into system design specifications.

4.2. Control Paradigms, Robustness, and Usability

Discrete versus continuous control. The reviewed systems fall into two control paradigms that have markedly different usability implications. The majority implement discrete command control, in which a recognized gesture is mapped to one of a small set of symbolic commands (e.g., forward, left, right, stop); this is simple, robust, and dominates the rule-based and most classification based studies (e.g., [3,18,19,22,38,41]). A smaller group implements continuous, proportional control, in which a continuously varying signal is mapped to continuously varying velocity or steering, for example the pressure to velocity and angular zone mapping of GestureMoRo by Chen et al. [20] and the proportional sEMG regression of Iqbal et al. [66]. Proportional control affords smoother, more natural motion and finer speed modulation but demands higher signal fidelity, careful calibration, and tighter latency budgets, and it is more sensitive to tremor and noise. This distinction is now annotated in the comparative tables and is an important, under discussed design axis: discrete control is generally safer for users with severe or fluctuating motor symptoms, whereas proportional control better serves users seeking naturalistic navigation in open environments.

Disadvantages of hybrid and multimodal designs. Although hybrid architectures (Tables 3 and 4) generally improve robustness, the review makes clear that this robustness is not free. Combining vision with inertial, pressure, or physiological sensing increases the bill of materials and power draw, multiplies the number of components that can fail or drift out of calibration, and complicates real-time synchronization of heterogeneous data streams [55,69]. Multimodal fusion raises computational load, which is at odds with low-cost embedded deployment, and each additional sensor adds a donning, placement, or maintenance burden for the user or caregiver. The integration of speech, physiological, and environmental sensing in systems such as that of Pasam et al. [69] also widens the privacy and data handling surface. Consequently, the "more modalities is better" intuition must be weighed against cost, energy, maintainability, and usability, constraints that are especially binding in low-resource settings.

Toward quantifiable environmental robustness testing. A recurring weakness across modalities is that environmental robustness is asserted qualitatively rather than measured. Vision studies note sensitivity to lighting and background, and sensor studies note drift, but few report performance as a function of controlled environmental variables. We, therefore, recommend that future evaluations adopt a standardized robustness protocol that sweeps and reports quantifiable conditions: illumination level (measured in lux, e.g., 50, 200, 500, and $>$1000 lux, plus backlit and outdoor sunlight), background clutter (uniform vs. cluttered vs. moving people backgrounds), user to camera distance and viewing angle, and, particularly relevant to tropical deployment, ambient temperature and relative humidity for sensor-based systems. Reporting accuracy, latency, and failure rate across such a grid would turn "robust under varying lighting" into a reproducible, comparable claim and would expose failure boundaries before clinical deployment.

User fatigue. Sustained gesture input imposes a physical and cognitive load that the reviewed literature rarely quantifies, yet which directly determines real-world adoption. Holding static poses (as required by several threshold-based vision systems [38,39]), performing repeated dynamic gestures, or maintaining sustained muscle contractions for sEMG control [9,14] can induce muscular fatigue, drift, and degraded recognition over a session, an effect compounded for users with neuromuscular conditions, for whom fatigue is itself a symptom. Fatigue also interacts with calibration: thresholds set on fresh muscle signals become inaccurate as the muscle tires. Future systems should minimize the frequency and effort of required gestures (e.g., latching commands, dwell free selection, and rest gestures), monitor signal quality to detect fatigue onset, and report session length usability rather than only short trial accuracy.

Individual calibration and personal adaptation. Because gesture morphology, tremor characteristics, electrode placement, and hand geometry vary substantially between users, fixed, population level models generalize poorly, a point made repeatedly across the reviewed studies (e.g., [5,9,14,37,54]). Several promising adaptation strategies appear in the literature and should be standard practice: per user threshold calibration from a short enrollment session [38,39]; normalization of landmark distances by hand size to achieve scale invariance [38]; transfer learning that adapts a base model to a new user with few samples [5,14]; and user specific validation mechanisms that detect and correct misclassifications arising from tremor, as in the two-stage validator of Bandara et al. [54], which improved Cohen's kappa from roughly 0.85–0.90 to about 0.98. The broader implication is that personalization should be treated as a first class design requirement rather than an afterthought, with lightweight on device adaptation that does not require cloud connectivity.

4.3. Cost Considerations for Low Resource Deployment

Cost is decisive for adoption in lower middle income settings, yet most studies report only model accuracy. To make the affordability argument concrete, Table 8 decomposes a gesture-controlled wheelchair into its major cost components and contrasts indicative low-cost (locally sourced, do it yourself) options against commercial or clinical equivalents. The figures are order of magnitude estimates compiled from the hardware described in the reviewed prototypes and from local market context; they are intended to illustrate relative cost structure rather than to constitute a formal costing study. Beyond one time hardware, the table highlights recurring and often ignored costs, batteries, calibration effort, maintenance, and technical support, which strongly affect total cost of ownership and sustainability where service networks are sparse.

Table 8. Indicative cost decomposition of a gesture-controlled wheelchair: low-cost (locally sourced) versus commercial/clinical options. Values are approximate USD ranges for illustration.

Component	Low-Cost Option (USD)	Commercial/Clinical (USD)	Notes
Controller/compute	ESP32/Raspberry Pi: 6–60	Embedded clinical controller: 300–800	Edge inference avoids cloud dependence
Camera (vision systems)	USB webcam: 8–25	Medical-grade/depth camera: 150–400	Webcam sufficient for MediaPipe landmarks
Motion/EMG sensor	MPU6050/low-cost EMG: 2–50	Clinical EMG/Myo-class: 200–1500	Dominant cost driver for sensor systems
Motors + drivers	2× DC motor + driver: 40–120	Clinical actuators: 400–1200	Often repurposed from a manual chair
Battery/power	Lead-acid/Li-ion pack: 30–120	Clinical battery system: 200–600	Recurring replacement cost
Frame/chassis	Repurposed manual chair: 50–200	Powered-chair frame: 800–3000	Reuse cuts cost substantially
Per-user calibration	Automated software: ≈0	Clinical fitting: 100–500	Favors on-device auto-calibration
Maintenance (annual)	Local parts: 20–60	Service contract: 200–800	Depends on local availability
Technical support	Community/local technician	OEM support contract	Sparse formal networks locally
Indicative system total	**≈200–700**	**≈1500–10,000**	Consistent with prices cited in Section 2

4.4. Positioning Relative to Existing Reviews

Several recent reviews survey hand gesture recognition, but from a technology centric rather than an application centric standpoint, and none targets gesture-based wheelchair navigation in low-resource settings. Cui et al. [70] review deep vision-based real-time hand gesture recognition; Ni et al. [71] survey surface electromyography based recognition; and Fertl et al. [72] review gesture recognition on edge devices, emphasizing sensor technologies, algorithms, and processing hardware. These works comprehensively cover algorithms and hardware for general gesture recognition, but they do not synthesize evidence around the assistive mobility use case, do not assess validation maturity or risk of bias for wheelchair systems, and do not consider the economic, environmental, and cultural constraints of deployment in countries such as Bangladesh. The present review is, therefore, complementary: it is organized around the wheelchair control application, integrates vision, sensor, and signal-based modalities within a single comparative and risk of bias framework (Tables 5–7), and explicitly foregrounds affordability, robustness, and cultural alignment. Looking forward, emerging illumination-invariant sensing hardware, such as printable meta infrared photodetector arrays [73], tunable mid infrared photodetectors based on graphene plasmonics [74], and heat assisted detection and ranging [75], points to a promising direction for gesture and obstacle sensing that is robust to the low-light and high glare conditions where conventional RGB vision fails, and merits investigation for next-generation assistive wheelchairs.

The economic reality of assistive technology in lower-middle-income countries demands fundamental rethinking of design priorities. Commercially available intelligent wheelchairs costing USD 1500–10,000 remain inaccessible to the vast majority of Bangladeshi users requiring mobility assistance, necessitating lightweight deep learning architectures compatible with embedded hardware costing under USD 100. Infrastructure constraints further compound technical requirements, as unstable electrical supply, limited internet connectivity, and absence of technical support networks necessitate edge-deployable systems with offline functionality and simplified maintenance requirements. The findings collectively argue against one-size-fits-all technological solutions, demonstrating that effective gesture-controlled wheelchairs must be contextualized within specific user populations, environmental conditions, cultural frameworks, and resource constraints. The Bangladeshi context presents unique challenges: high population density creating crowded navigation environments, tropical climate with extreme humidity affecting sensor performance, linguistic diversity requiring culturally aligned gesture vocabularies, and economic constraints demanding affordable, maintainable solutions.

Based on these findings, we developed a context-aware framework specifically tailored to Bangladesh that addresses identified limitations through four integrated components. Lightweight deep learning models optimized for edge deployment on low-cost embedded hardware incorporate model compression techniques including depthwise separable convolutions, knowledge distillation, and pruning to enable deployment on resource-constrained platforms. Hybrid vision-sensor architectures combine camera-based MediaPipe landmark extraction with inertial measurement units, providing complementary robustness and redundancy when visual conditions degrade. Gesture compensation mechanisms accommodate motor impairments including pathological tremor through adaptive threshold mechanisms that accommodate user-specific tremor characteristics via personalized calibration. Culturally aligned gesture vocabularies derived from Bangladeshi social conventions engage local stakeholders to identify intuitively meaningful hand configurations aligned with regional non-verbal communication norms. This framework prioritizes computational efficiency without sacrificing clinical utility, demonstrating how systematic evidence synthesis can inform culturally grounded, economically viable, and technically robust assistive technology development.

The following subsection translates the synthesis, validation profile (Tables 6 and 7), and cost analysis (Table 8) of this review into a prioritized, evidence-based research agenda. We define future research directions as the set of concrete investigations needed to move gesture-controlled wheelchairs from laboratory prototypes toward affordable, robust, and clinically relevant assistive devices deployable in low-resource environments such as Bangladesh.

4.5. Future Research Directions

The future research agenda should be directly derived from the limitations identified in this review rather than framed as broad technological improvement. The first priority is to close the gap between laboratory accuracy and clinically usable wheelchair control. The validation profile of the reviewed studies shows that most systems remain offline or prototype-level, with limited testing involving motor-impaired, elderly, or real wheelchair users. Therefore, future studies should move from image- or signal-classification experiments to complete wheelchair navigation trials. These trials should evaluate route completion, command accuracy, response latency, collision events, emergency-stop activation, user workload, and per-user failure cases. To make these studies reproducible, authors should publish the route layout, obstacle configuration, participant inclusion criteria, sensor placement, calibration procedure, raw and processed logs, train-test splits, source code, and model weights.

The second priority is quantifiable environmental robustness. This review identified illumination variation, background clutter, occlusion, z-axis instability, ambient light interference, sensor drift, humidity, and noise as recurring causes of system failure across vision-based, sensor-based, and multimodal systems. Future evaluations should, therefore, include a standardized environmental stress protocol rather than only stating that the system works under "different conditions." For vision-based systems, performance should be reported under controlled lux levels, backlighting, outdoor sunlight, cluttered backgrounds, moving-person backgrounds, different user-camera distances, and different viewing angles. For wearable, EMG, IMU, optical, and pressure-based systems, performance should also be tested under temperature and humidity variation. Each condition should report accuracy, latency, false command rate, missed command rate, and failure boundary so that robustness claims can be independently reproduced.

The third priority is user-specific adaptation for pathological tremor, fatigue, and motor variability. The review shows that only a small fraction of studies explicitly addressed

tremor or unintended movement, even though the target users of smart wheelchairs often experience inconsistent motor patterns. Future systems should not rely only on fixed thresholds or population-level models. They should include short per-user calibration, tremor-aware validation, adaptive thresholds, continual learning, and fatigue-aware signal monitoring. Long-session testing is needed because gesture quality, EMG signals, pressure input, and head or hand movement can change as the user becomes tired. Reproducible reporting should include tremor severity, fatigue measurement, session duration, calibration time, per-user confusion matrices, and whether the model was tested within-user or between-user.

The fourth priority is cost-aware multimodal fusion. The cross-modality synthesis shows that no single input modality satisfies all requirements: vision is contactless but sensitive to lighting and occlusion, while wearable and physiological sensors are less affected by lighting but require calibration, placement, and maintenance. Future work should, therefore, examine adaptive fusion, where one modality acts as the primary command channel and another modality provides backup, validation, or safety override when failure conditions occur. However, fusion should not simply add more sensors. Each added sensor should be justified by measurable improvement in safety, robustness, or usability. Reproducibility requires ablation results for vision-only, sensor-only, and fused configurations, together with sensor synchronization methods, latency, power consumption, and total hardware cost.

The fifth priority is edge-deployable and maintainable design for low-resource environments. Since the review specifically targets Bangladesh and similar settings, future systems should demonstrate operation on affordable embedded hardware with offline functionality. Studies should report model size, memory use, inference speed, battery runtime, maintenance requirements, and bill of materials. Model compression, pruning, quantization, and lightweight architectures should be evaluated not only by recognition accuracy but also by whether they can run reliably on low-cost controllers without cloud dependence. Reproducibility requires open firmware, hardware diagrams, component lists, and deployment instructions that allow another team to rebuild the system.

The sixth priority is to move from gesture recognition alone to complete intelligent wheelchair navigation. Future systems should integrate user intention recognition with obstacle detection, safe command execution, navigation planning, health or emergency monitoring where appropriate, and explainable decision feedback. A gesture should not be evaluated only as a correct label; it should be evaluated by whether it produces safe and acceptable wheelchair movement in real environments. Reproducible studies should, therefore, define standard indoor and outdoor test routes, obstacle layouts, safety rules, failure definitions, and decision logs.

Finally, future research should develop culturally aligned gesture vocabularies for Bangladeshi and South Asian users. The review identified cultural mismatch as an unresolved limitation because many gesture sets are borrowed from general datasets or foreign sign-language conventions. Future work should co-design gesture commands with wheelchair users, caregivers, physiotherapists, and local technicians, then test whether those gestures are intuitive, physically comfortable, distinguishable, and acceptable in local social contexts. To make this work reproducible, researchers should publish the gesture dictionary, video examples, annotation protocol, participant demographics, confusion matrices, and user acceptability results. This would ensure that future systems are not only technically accurate but also clinically usable, culturally meaningful, affordable, and independently verifiable.

5. Conclusions

Gesture-controlled intelligent wheelchairs represent a promising direction for assistive mobility technologies; however, current research remains largely constrained to controlled laboratory environments and often overlooks real-world usability challenges. This systematic review highlights critical gaps in existing literature, including limited validation across diverse user populations, insufficient consideration of pathological motor impairments, lack of environmental robustness, and the absence of culturally relevant gesture vocabularies. Addressing these limitations is particularly important for low-resource contexts such as Bangladesh, where economic, infrastructural, and social factors strongly influence technology adoption.

This review highlights potential strategies for addressing these challenges in the Bangladesh context, including lightweight deep learning models suitable for edge devices, hybrid vision–sensor architectures for improved reliability, gesture compensation mechanisms for motor impairments, and culturally aligned gesture vocabularies tailored to local users. Future research should focus on large-scale real-world validation with diverse users, longitudinal studies examining long-term usability and adaptation, standardized evaluation protocols for fair comparison across studies, and comprehensive cost and sustainability analyses for deployment in low-resource environments. By prioritizing real-world usability and user-centered design, future developments in gesture-controlled wheelchair systems can move beyond laboratory prototypes toward practical assistive technologies that meaningfully enhance independent mobility for people with disabilities.

Author Contributions: Conceptualization, R.A.D. and S.H.C.; methodology, R.A.D. and M.A.A.M.; formal analysis, R.A.D., M.S.H. and S.B.; investigation, R.A.D. and M.S.R.; resources, M.A.K.A.; data curation, R.A.D. and S.H.C.; writing—original draft preparation, R.A.D.; writing—review and editing, R.A.D., S.H.C. and S.B.; visualization, R.A.D. and M.S.H.; supervision, M.A.K.A.; project administration, M.A.K.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was conducted under the Improving Computer and Software Engineering Tertiary Education Project (ICSETEP), Subproject D-67, titled "Developing a Gesture-Based Intelligent Wheelchair for Assisting the Physically Challenged People in Bangladesh", implemented by the Department of Computer Science and Engineering, Jahangirnagar University. The project was funded by the Asian Development Bank (ADB) and the Government of Bangladesh (GoB).

Institutional Review Board Statement: Not applicable. This study is a review paper and did not involve humans or animals.

Informed Consent Statement: Not applicable. This study did not involve humans.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors would like to acknowledge the support of the ICSETEP project and the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. During the preparation of this manuscript, the authors used ChatGPT (GPT-5.5) for language refinement and manuscript editing purposes. The authors reviewed and edited the generated content and take full responsibility for the final published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AE	Absolute Envelope
AHC	Agglomerative Hierarchical Clustering
AI	Artificial Intelligence
ANN	Artificial Neural Network
BiLSTM	Bidirectional Long Short-Term Memory
CNN	Convolutional Neural Network
DASDV	Difference Absolute Standard Deviation Value
DL	Deep Learning
DQN	Deep Q-Network
EMG	Electromyography
EWL	Enhanced Waveform Length
FIR	Finite Impulse Response
GA	Genetic Algorithm
GRU	Gated Recurrent Unit
HMI	Human–Machine Interface
IMU	Inertial Measurement Unit
IoMT	Internet of Medical Things
KCF	Kernelized Correlation Filter
KNNs	K-Nearest Neighbors
LD	Log Detector
LSTM	Long Short-Term Memory
MAR	Mouth Aspect Ratio
MAV	Mean Absolute Value
MDP	Markov Decision Process
ML	Machine Learning
MLP	Multi-Layer Perceptron
MYOP	Myopulse Percentage Rate
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSO	Particle Swarm Optimization
RMS	Root Mean Square
RNN	Recurrent Neural Network
RoB	Risk of Bias
SD	Standard Deviation
sEMG	Surface Electromyography
SVM	Support Vector Machine
VAR	Variance
VLC	Visible Light Communication
WL	Waveform Length
YOLO	You Only Look Once
ZC	Zero Crossings

References

1. World Health Organization; World Bank. *World Report on Disability 2011*; World Health Organization: Geneva, Switzerland, 2011.
2. Keeler, E.; Guralnik, J.M.; Tian, H.; Wallace, R.B.; Reuben, D.B. The impact of functional status on life expectancy in older persons. *J. Gerontol. Ser. Biomed. Sci. Med. Sci.* **2010**, *65*, 727–733. [CrossRef]
3. Islam, M.A.; Hasib, A.; Emon, T.M.; Hasan, K.T. An AI-IoT Based Smart Wheelchair with Gesture-Controlled Mobility, Deep Learning-Based Obstacle Detection, Multi-Sensor Health Monitoring, and Emergency Alert System. In *Proceedings of the 2025 IEEE 2nd International Conference on Computing, Applications and Systems (COMPAS)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6.
4. Basak, S.; Nandiny, F.F.; Chowdhury, S.M.H.; Biswas, A.A. Gesture-based smart wheelchair for assisting physically challenged people. In *Proceedings of the 2021 International Conference on Computer Communication and Informatics (ICCCI)*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.

5. Vásconez, J.P.; Barona López, L.I.; Valdivieso Caraguay, Á.L.; Benalcázar, M.E. Hand gesture recognition using EMG-IMU signals and deep q-networks. *Sensors* **2022**, *22*, 9613. [CrossRef] [PubMed]
6. Nogales, R.E.; Benalcázar, M.E. Hand gesture recognition using automatic feature extraction and deep learning algorithms with memory. *Big Data Cogn. Comput.* **2023**, *7*, 102. [CrossRef]
7. Wang, S.; Zhang, T.; Li, Y.; Li, P.; Wu, H.; Li, K. Continuous hand gestures detection and recognition in emergency human-robot interaction based on the inertial measurement unit. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2526315. [CrossRef]
8. Zhao, J.; Li, X.H.; Cruz, J.C.D.; Verdadero, M.S.; Centeno, J.C.; Novelero, J.M. Hand gesture recognition based on deep learning. In *Proceedings of the 2023 International Conference on Digital Applications, Transformation & Economy (ICDATE)*; IEEE: Piscataway, NJ, USA, 2023; pp. 250–254. [CrossRef]
9. Bao, T.; Lu, Z.; Zhou, P. Deep learning-based post-stroke myoelectric gesture recognition: From feature construction to network design. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2024**, *33*, 191–200.
10. Zhang, Z.; Shen, Q.; Wang, Y. Electromyographic hand gesture recognition using convolutional neural network with multi-attention. *Biomed. Signal Process. Control* **2024**, *91*, 105935. [CrossRef]
11. Fernandez, F.D.M.; Kim, M.; Yoon, S.; Kim, J. Capacitive BaTiO3-PDMS hand-gesture sensor: Insights into sensing mechanisms and signal classification with machine learning. *Compos. Sci. Technol.* **2024**, *251*, 110581. [CrossRef]
12. Kateb, P.; Fornaciari, A.; Ahmadizadeh, C.; Shokurov, A.; Cicoira, F.; Menon, C. High-Performance Textile-Based Capacitive Strain Sensors via Enhanced Vapor Phase Polymerization of Pyrrole and Their Application to Machine Learning-Assisted Hand Gesture Recognition. *Adv. Intell. Syst.* **2024**, *6*, 2400292. [CrossRef]
13. Kaur, G.; Gupta, B.; Saini, D.; Mohindru, K.; Nehra, A. SecGes: A Secure Biometric-Driven Hand Gesture Recognition System for Smart Wheelchair Movement. In *Proceedings of the 2023 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
14. Zhang, Y.; Chen, Y.; Yu, H.; Yang, X.; Lu, W. Dual layer transfer learning for sEMG-based user-independent gesture recognition. *Pers. Ubiquitous Comput.* **2022**, *26*, 575–586.
15. Rusydi, M.I.; Sari, Y.A.R.; Jordan, A.; Syafii; Rahmadani, F.; Nugroho, H. Hand Gesture Wheelchair Control Based on the Wrist Rotation with Agglomerative Hierarchical Clustering Method. In *Proceedings of the 2023 8th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*; IEEE: Piscataway, NJ, USA, 2023; pp. 95–99.
16. Singh, G.; Gautam, S.; Puri, Y.; Parikh, J.; Mishra, G. Novel system based on amalgamation of embedded systems and edge ai for hand gesture recognition: An edge ai solution for hand gesture recognition. In *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing*; ACM: New York, NY, USA, 2024; pp. 432–439.
17. Anam, K.; Nahela, S.; Sasono, M.A.H.; Rizal, N.A.; Putra, A.N.; Wahono, B.; Putrasari, Y.; Wardana, M.K.A.; Salim, T.I. Electric wheelchair navigation based on hand gestures prediction using the k-Nearest Neighbor method. *J. Mechatron. Electr. Power Veh. Technol.* **2025**, *16*, 132–143. [CrossRef]
18. Mahdin, M.E.; Ornob, T.M.; Akhter, M.E.; Mullick, S.A.; Arnob, S.S.; Palit, R.; Habib, A.B. Hand gesture-controlled smart wheelchair with gps tracking and deep learning-based object detection for collision avoidance. In P*roceedings of the 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*; IEEE: Piscataway, NJ, USA, 2022; pp. 147–152.
19. Calado, A.; Roselli, P.; Errico, V.; Magrofuoco, N.; Vanderdonckt, J.; Saggio, G. A geometric model-based approach to hand gesture recognition. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**, *52*, 6151–6161. [CrossRef]
20. Chen, L.; Li, C.; Fahmy, A.; Sienz, J. GestureMoRo: An algorithm for autonomous mobile robot teleoperation based on gesture recognition. *Sci. Rep.* **2024**, *14*, 6199. [CrossRef] [PubMed]
21. Rambabu, K.; Teju, E.; Nandini, G.; Pooja, K. Smart Wheelchair with Hand Gesture Control and Real-Time Health Monitoring Using IoT. In *Proceedings of the 2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1937–1945.
22. Balaji, S.; Sree, R.V.; Elma, K.J. Smart Wheelchair based on Gesture Control with Autonomous Navigation. In *Proceedings of the 2025 7th International Conference on Inventive Material Science and Applications (ICIMA)*; IEEE: Piscataway, NJ, USA, 2025; pp. 803–811.
23. Gopal, P.; Gesta, A.; Mohebbi, A. A systematic study on electromyography-based hand gesture recognition for assistive robots using deep learning and machine learning models. *Sensors* **2022**, *22*, 3650. [CrossRef] [PubMed]
24. Gadekallu, T.R.; Srivastava, G.; Liyanage, M.; M., I.; Chowdhary, C.L.; Koppu, S.; Maddikunta, P.K.R. Hand gesture recognition based on a Harris hawks optimized convolution neural network. *Comput. Electr. Eng.* **2022**, *100*, 107836. [CrossRef]
25. Bhushan, S.; Alshehri, M.; Keshta, I.; Chakraverti, A.K.; Rajpurohit, J.; Abugabah, A. An experimental analysis of various machine learning algorithms for hand gesture recognition. *Electronics* **2022**, *11*, 968. [CrossRef]
26. Sadi, M.S.; Alotaibi, M.; Islam, M.R.; Islam, M.S.; Alhmiedat, T.; Bassfar, Z. Finger-gesture-controlled wheelchair with enabling IoT. *Sensors* **2022**, *22*, 8716. [CrossRef] [PubMed]

27. Sahoo, J.P.; Prakash, A.J.; Pławiak, P.; Samantray, S. Real-time hand gesture recognition using fine-tuned convolutional neural network. *Sensors* **2022**, *22*, 706. [CrossRef] [PubMed]
28. Zhou, W.; Chen, K. A lightweight hand gesture recognition in complex backgrounds. *Displays* **2022**, *74*, 102226. [CrossRef]
29. Dang, T.L.; Tran, S.D.; Nguyen, T.H.; Kim, S.; Monet, N. An improved hand gesture recognition system using keypoints and hand bounding boxes. *Array* **2022**, *16*, 100251. [CrossRef]
30. Padhi, P.; Das, M. Hand gesture recognition using densenet201-mediapipe hybrid modelling. In *Proceedings of the 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*; IEEE: Piscataway, NJ, USA, 2022; pp. 995–999.
31. Mohamed, T.; Ibrahim, A.; Faiz, T.; Alhasan, W.; Atta, A.; Mago, V.; Ejaz, M.A.; Munir, S. Intelligent hand gesture recognition system empowered with CNN. In *Proceedings of the 2022 International Conference on Cyber Resilience (ICCR)*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–8.
32. Wu, S.; Li, Z.; Li, S.; Liu, Q.; Wu, W. Static gesture recognition algorithm based on improved YOLOv5s. *Electronics* **2023**, *12*, 596. [CrossRef]
33. Jafari, F.; Basu, A. Two-dimensional parallel spatio-temporal pyramid pooling for hand gesture recognition. *IEEE Access* **2023**, *11*, 133755–133766. [CrossRef]
34. Kumar, A.; Tomar, R.; Raghav, D.; Yadav, A. Patient-Centric Hand-Gesture Driven Wheelchair Control Using Convolutional Neural Networks for Enhanced Healthcare Mobility. In *Proceedings of the 2025 3rd International Conference on Disruptive Technologies (ICDT)*; IEEE: Piscataway, NJ, USA, 2025; pp. 978–983.
35. Do Tran, V.; Nguyen, C.T. Design of a Wheelchair Control System Based on Hand Gesture Recognition Using ResNet18. *J. Tech. Educ. Sci.* **2026**. [CrossRef]
36. Nivash, S.; Ganesh, E.; Manikandan, T.; Dhaka, A.; Nandal, A.; Hoang, V.T.; Kumar, A.; Belay, A. Implementation and Analysis of AI-Based Gesticulation Control for Impaired People. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 4656939. [CrossRef]
37. Khaksar, S.; Checker, L.; Borazjan, B.; Murray, I. Design and evaluation of an alternative control for a quad-rotor drone using hand-gesture recognition. *Sensors* **2023**, *23*, 5462. [CrossRef] [PubMed]
38. Huda, M.R.; Ali, M.L.; Sadi, M.S. Real-time hand-gesture recognition for the control of wheelchair. In *Proceedings of the 2022 12th International Conference on Electrical and Computer Engineering (ICECE)*; IEEE: Piscataway, NJ, USA, 2022; pp. 384–387.
39. Huda, M.R.; Ali, M.L.; Sadi, M.S. Developing a real-time hand-gesture recognition technique for wheelchair control. *PLoS ONE* **2025**, *20*, e0319996. [CrossRef] [PubMed]
40. Drăgoi, M.V.; Nisipeanu, I.; Puiu, R.A.; Tache, F.G.; Spiridon-Mocioacă, T.M.; Hank, A.; Cristoiu, C. An Inclusive Offline Learning Platform Integrating Gesture Recognition and Local AI Models. *Biomimetics* **2025**, *10*, 693. [CrossRef] [PubMed]
41. Ritu, K.J.; Mahmud, S.S.; Bhuiyan, S.; Ohidujjaman; Hasan, M. A Facial Gesture-Controlled Wheelchair for Individuals with Complete Disabilities. In *Proceedings of the 2025 2nd International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6.
42. D'Souza, L.; Kulal, B.G.; Shrinivas, U.; Shetty, P.M.; Singh, C.; Bekal, A. Design & implementation of novel AI based hand gestured smart wheelchair. In *Proceedings of the 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
43. Abiraj, B.; Sethuraman, K.; Bethanney Janney, J.; Grace Kanmani Prince, P.; Sudhakar, T.; Ezekiel Jebaraj, S. AI-powered automated wheelchair with lane detection. In *Proceedings of the 2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–5.
44. He, D.; Yang, Y.; Wu, R. Design of human–computer interaction gesture tracking model-based on improved pso and kcf algorithms. *IEEE Access* **2024**, *12*, 39913–39925. [CrossRef]
45. Leon, D.G.; Gröli, J.; Yeduri, S.R.; Rossier, D.; Mosqueron, R.; Pandey, O.J.; Cenkeramaddi, L.R. Video hand gestures recognition using depth camera and lightweight CNN. *IEEE Sens. J.* **2022**, *22*, 14610–14619. [CrossRef]
46. Peral, M.; Sanfeliu, A.; Garrell, A. Efficient hand gesture recognition for human-robot interaction. *IEEE Robot. Autom. Lett.* **2022**, *7*, 10272–10279. [CrossRef]
47. Rehman, M.U.; Ahmed, F.; Khan, M.A.; Tariq, U.; Alfouzan, F.A.; Alzahrani, N.M.; Ahmad, J. Dynamic hand gesture recognition using 3D-CNN and LSTM networks. *Comput. Mater. Contin.* **2022**, *70*, 4675. [CrossRef]
48. Nguyen, T.T.; Nguyen, N.C.; Ngo, D.K.; Phan, V.L.; Pham, M.H.; Nguyen, D.A.; Doan, M.H.; Le, T.L. A continuous real-time hand gesture recognition method based on skeleton. In *Proceedings of the 2022 11th International Conference on Control, Automation and Information Sciences (ICCAIS)*; IEEE: Piscataway, NJ, USA, 2022; pp. 273–278.
49. Miah, A.S.M.; Hasan, M.A.M.; Shin, J. Dynamic hand gesture recognition using multi-branch attention based graph and general deep learning model. *IEEE Access* **2023**, *11*, 4703–4716. [CrossRef]
50. Mohammed, A.A.; Lv, J.; Islam, M.S.; Sang, Y. Multi-model ensemble gesture recognition network for high-accuracy dynamic hand gesture recognition. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 6829–6842.
51. Narayanan, A.L.; Wu, H.; Ma, J.; Huang, M. RADFormers for detection and classification of hand gestures. *IEEE Sens. J.* **2024**, *24*, 9093–9103. [CrossRef]

52. Wang, Y.; Jia, X.; Zhou, M.; Xie, L.; Tian, Z. A novel F-RCNN based hand gesture detection approach for FMCW systems. *Wirel. Netw.* **2024**, *30*, 3395–3408.
53. Bremer, G.; McIntyre, J.; Jung, J.H.; Ellero, S.; Mouawad, I.; Di Gloria, D.; Lappe, M. A Bi-Directional Deep Learning Interface for Gaze-Controlled Wheelchair Navigation: Overcoming the Midas Touch Problem. In *Proceedings of the 2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*; IEEE: Piscataway, NJ, USA, 2025; pp. 867–877.
54. Bandara, H.R.T.; Priyanayana, K.; Rajendran, H.; Chandima, D.; Jayasekara, A.B.P. Hand gesture classification model for intelligent wheelchair with improved gesture variance compensation. In *Proceedings of the 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*; IEEE: Piscataway, NJ, USA, 2023; pp. 4708–4713.
55. Cui, J.; Shang, Y.; Yu, S.; Wang, Y. Research on intelligent wheelchair multimode human–computer interaction and assisted driving technology. *Actuators* **2024**, *13*, 230. [CrossRef]
56. Chamalsha, A.; Nadeeshan, R.; Herath, H.; Yasakethu, S.; Oruthota, U. User-Centric Design of Contactless Smart Wheelchair System for Elderly People. In *Proceedings of the 2024 6th International Conference on Advancements in Computing (ICAC)*; IEEE: Piscataway, NJ, USA, 2024; pp. 456–461.
57. Sirisati, R.S.; Eenaja, A.; Sreeja, N.; Kumar, T.S.; Naresh, B. Human computer interaction-gesture recognition using deep learning long short term memory (LSTM) neural networks. *J. Next Gener. Technol.* **2024**, *4*, 9–18.
58. Meghna, S.; VM, S.S.; Bhat, A.; Kharvi, R.K.; Satrajeeth, S. Simulation-Based Smart Wheelchair Navigation Using YOLO and R-CNN with Customized A* Algorithm. In *Proceedings of the 2025 Control Instrumentation System Conference (CISCON)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6.
59. Bhawarthi, A.; Lokhande, N.; Lokhande, S.; Lole, R.; Lonkar, T.; Loya, V.; Lunawat, B. Hand Gesture Controlled Wheelchair Using Image Processing. In *Proceedings of the 2024 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
60. Madaan, V.; Sharma, N.; Gupta, D.; Aluvala, S. Hand Gesture Detection Using VGG-16. In *Proceedings of the 2024 International Conference on Information Science and Communications Technologies (ICISCT)*; IEEE: Piscataway, NJ, USA, 2024; pp. 491–495.
61. Awaluddin, B.A.; Chao, C.T.; Chiou, J.S. A hybrid image augmentation technique for user-and environment-independent hand gesture recognition based on deep learning. *Mathematics* **2024**, *12*, 1393. [CrossRef]
62. Patankar, N.S.; Shaharam, D.A.; Rajendra, D.T.; Bhausaheb, G.S.; Satish, P.S.; Deshmukh, Y.S. IoT Based Hand Gesture Control Wheelchair for Disabled Person. In *Proceedings of the 2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*; IEEE: Piscataway, NJ, USA, 2023; pp. 346–352.
63. Kalantri, R.A.; Chitre, D. Automatic wheelchair using gesture recognition. *Int. J. Eng. Innov. Technol. (IJEIT)* **2013**, *2*, 216–218.
64. Yang, W.; Liu, S.; Wang, Z.; Liu, H.; Pan, C.; Liu, C.; Shen, C. Bioinspired composite fiber aerogel pressure sensor for machine-learning-assisted human activity and gesture recognition. *Nano Energy* **2024**, *127*, 109799. [CrossRef]
65. Iqbal, H.; Zheng, J.; Chai, R.; Chandrasekaran, S. K-Nearest Neighbours and Ensemble Based Real-Time Hand Gesture Recognition for Powered Wheelchair. In *Proceedings of the 2023 International Conference on Advanced Mechatronic Systems (ICAMechS)*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
66. Iqbal, H.; Zheng, J.; Chai, R.; Chandrasekaran, S. Regression-based real-time hand gesture recognition and control for electric powered wheelchair. In Proceedings of the Australasian Conference on Robotics and Automation (ACRA 2022), Brisbane, Australia, 6–8 December 2022.
67. Nithya, B.; Sanjay, S.; Thanush, R.; Thejas, B.; Venugeetha, Y. Wheelchair Control and Home Automation Using Hand Gestures. *EAI Endorsed Trans. Smart Cities* **2022**, *4*, e3.
68. Liang, X.; Li, J.; Xu, C.; Xie, Z.; Zhang, C.; Ding, W.; Gui, W. Gesture recognition and control based on visible light communication. In *Proceedings of the 2024 43rd Chinese Control Conference (CCC)*; IEEE: Piscataway, NJ, USA, 2024; pp. 6164–6169.
69. Pasam, G.; Solano, D.C.; Arampattammel, J.P.; AL-Habsi, M.H.; AI-Nudhiri, A.Z.A.; Husaini, S.M.A. Solar-Powered Autonomous Robotic Wheelchair with AI, IVRS, Facial Recognition, and Multi-Tasking Capabilities. In *Proceedings of the 2025 IEEE 6th International Conference in Robotics and Manufacturing Automation (ROMA)*; IEEE: Piscataway, NJ, USA, 2025; pp. 292–297.
70. Cui, C.; Sunar, M.S.; Eg Su, G. Deep vision-based real-time hand gesture recognition: A review. *PeerJ Comput. Sci.* **2025**, *11*, e2921. [CrossRef] [PubMed]
71. Ni, S.; Al-qaness, M.A.A.; Hawbani, A.; Al-Alimi, D.; Abd Elaziz, M.; Ewees, A.A. A survey on hand gesture recognition based on surface electromyography: Fundamentals, methods, applications, challenges and future trends. *Appl. Soft Comput.* **2024**, *166*, 112235. [CrossRef]
72. Fertl, E.; Castillo, E.; Stettinger, G.; Cuéllar, M.P.; Morales, D.P. Hand Gesture Recognition on Edge Devices: Sensor Technologies, Algorithms, and Processing Hardware. *Sensors* **2025**, *25*, 1687. [CrossRef] [PubMed]
73. Guo, J.; Gu, S.; Lin, L.; Liu, Y.; Cai, J.; Cai, H.; Tian, Y.; Zhang, Y.; Zhang, Q.; Liu, Z.; et al. Type-printable photodetector arrays for multichannel meta-infrared imaging. *Nat. Commun.* **2024**, *15*, 5193. [CrossRef] [PubMed]

74. Guo, J.; Lin, L.; Li, S.; Chen, J.; Wang, S.; Wu, W.; Cai, J.; Zhou, T.; Liu, Y.; Huang, W. Ferroelectric superdomain controlled graphene plasmon for tunable mid-infrared photodetector with dual-band spectral selectivity. *Carbon* **2022**, *189*, 596–603. [CrossRef]
75. Bao, F.; Wang, X.; Sureshbabu, S.H.; Sreekumar, G.; Yang, L.; Aggarwal, V.; Boddeti, V.N.; Jacob, Z. Heat-assisted detection and ranging. *Nature* **2023**, *619*, 743–748. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.