Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

# Development of a Foundation AI-Based Multimodal Human Machine Interaction Framework for Intelligent Accelerometer-Assisted Robotic Hand Glove Control in Parkinson's Disease

**Princy Naveena Chandrasegar**
Karunya Institute of Technology and Sciences

**Rajalakshmy Pallissery**
Karunya Institute of Technology and Sciences

**Bondala Ramakrishna**
Aditya Institute of Technology and Management

**Vairavel Madheswaran**
vairavel.m@dsce.ac.in

Dhanalakshmi Srinivasan College of Engineeri

**Rajpradesh Thangaraj**
Kalasalingam Academy of Research and Education

**Article**

**Keywords:** Parkinson's Disease, Foundation Artificial Intelligence, Robotic Hand Glove, Accelerometer-Assisted Motion Control, Vision-Based Hand Gesture Recognition, Multimodal Foundation Transformer, Human–Machine Interaction

**Posted Date:** October 16th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10858203/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

**Additional Declarations:** No competing interests reported.

# DEVELOPMENT OF A FOUNDATION AI-BASED MULTIMODAL HUMAN MACHINE INTERACTION FRAMEWORK FOR INTELLIGENT ACCELEROMETER-ASSISTED ROBOTIC HAND GLOVE CONTROL IN PARKINSON'S DISEASE

**Princy Naveena Chandrasegar[1], Rajalakshmy Pallissery[2], Bondala Ramakrishna[3], Vairavel Madeshwaren[4*], Rajpradeesh Thangaraj[5]**

[1*] Research scholar, Karunya Institute of Technology and Sciences, Coimbatore - 641114, India.

Email: princyvc@karunya.edu.in

[2]Associate Professor, Karunya Institute of Technology and Sciences, Coimbatore - 641114, India

Email: rajalakshmy@karunya.edu.in

[3]Department of Mechanical Engineering, Aditya Institute of Technology and Management, Tekkali–532203, India

Email: brkbtech@gmail.com

[4*]Professor, General Engineering Dhanalakshmi Srinivasan College of Engineering,

Coimbatore,TamilNadu, India – 600 034

phdannauniv2020@gmail.com

[5]Department of Mechanical Engineering,

Kalasalingam Academy of Research and Education, Krishnankoil–626126, Tamil Nadu, India

Email: pradeeshr132@gmail.com

Corresponding mail id: vairavel.m@dsce.ac.in

**Acknowledgement**

The authors would like to express their sincere gratitude to all individuals and organizations that made their publicly available resources and data accessible, which supported the completion of this research. The authors also appreciate the valuable comments and suggestions provided by the reviewers and editors, which helped improve the quality of this manuscript.

**Abstract**

This study proposes a Foundation Artificial Intelligence (Foundation AI)-based multimodal human–machine interaction framework for an intelligent accelerometer-assisted robotic hand glove to support personalized upper-limb rehabilitation in Parkinson's disease patients. Approximately 150 participants (120 Parkinson's disease patients with Hoehn and Yahr Stage II–III and 30 healthy elderly volunteers) perform six clinically relevant Activities of Daily Living (ADLs), including grasping, eating, drinking, writing, dressing, and opening doors, to capture realistic neuromuscular movement patterns. The lightweight soft robotic glove integrates surface EMG sensors, an IMU incorporating a three-axis accelerometer and gyroscope, flex sensors, force-sensitive resistors, pressure sensors, vibration sensors, and temperature sensors to continuously monitor muscle activity, hand kinematics, tremor characteristics, grip force, wrist orientation, and movement acceleration. The embedded accelerometer enhances movement intention recognition by detecting motion direction, velocity, posture transitions, and dynamic hand activities. To improve accessibility, the framework incorporates three intelligent human–machine interaction modules: voice command control using Whisper and a Bidirectional Transformer for speech-driven robotic assistance; vision-based control via an external webcam, employing MediaPipe Hands for real-time hand landmark detection and Vision Transformer (ViT) for gesture recognition and movement intention estimation; and a mobile healthcare platform supporting personalized rehabilitation, remote monitoring, and secure telemedicine communication through Bluetooth Low Energy (BLE), MQTT, and AES-256 encryption. The visual features extracted from the webcam are fused with multimodal physiological and motion signals through a Multimodal Foundation Transformer (MFT) incorporating cross-modal attention and self-supervised contrastive

learning to estimate movement intention, tremor severity, hand gesture, finger trajectory, grasp type, and grip force. An Adaptive Impedance Controller subsequently personalizes robotic assistance by regulating actuator torque, finger motion, and wrist orientation in real time while continuously adapting to patient-specific neuromuscular conditions. The proposed framework establishes an intelligent multimodal perception and control architecture for personalized robotic rehabilitation and provides the computational foundation for future Digital Twin-enabled adaptive neuro-rehabilitation.

**Keywords:** Parkinson's Disease; Foundation Artificial Intelligence; Robotic Hand Glove; Accelerometer-Assisted Motion Control; Vision-Based Hand Gesture Recognition; Multimodal Foundation Transformer; Human–Machine Interaction.

## 1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder, whose motor symptoms, mainly resting tremor, bradykinesia, rigidity and postural instability, greatly impair hand dexterity and Activities of Daily Living (ADLs). As the disease progresses, patients become more and more unable to perform daily activities such as grasping, writing, eating, dressing, and using objects, and lose their functional independence and quality of life. Very recent developments in wearable robotics, artificial intelligence and multimodal sensing have opened up new possibilities for intelligent rehabilitation systems that can continuously monitor neuromuscular activity and provide personalized robotic support. However, current rehabilitation methods mostly use unimodal sensing methods and traditional machine learning methods, which fail to provide precise understanding of human intentions and adjust the robot's response in real-time. However, these challenges inspire this study to develop an intelligent accelerometer-assisted robotic hand glove with multimodal human–machine interaction framework based on the foundation of AI to achieve personalized and intelligent rehabilitation for patients with Parkinson's disease, using both physiological sensing and vision, speech interaction, multimodal transformer learning, adaptive impedance control and IoT-enabled mobile healthcare.

### 1.1 Problem Statement

Although significant advances have been made in the field of robotic rehabilitation technologies, current intelligent hand rehabilitation systems are still limited in their ability of multi-modal perception, movement intention recognition, adaptability to patient specific motor impairments, and real-time human–machine interaction. Today's systems are largely based on individual physiological sensors or vision-based methods which are unable to combine all the heterogeneous information from the neuromuscular, motion, visual and speech signals to make complete rehabilitation decisions. Moreover, traditional deep learning models need large amounts of labeled data and have little generalization ability for different rehabilitation conditions. Therefore, it is important to develop a multimodal intelligent rehabilitation system that can have a strong movement intention estimation, tremor severity assessment, adaptive control of robotic gloves, and personalized rehabilitation using Foundation Artificial Intelligence and multimodal representation learning.

## 2. Related Work

The development of wearable rehabilitation technologies has revolutionized the way Parkinson's disease is treated and managed, and these technologies offer ongoing treatment outside of the clinical setting [1]. Smart glove rehabilitation systems mostly targeted repetitive hand exercises and motor recovery with sensor-based rehabilitation protocols [2] to date. Due to the growing incidence of Parkinson's disease, researchers have designed smart rehabilitation gloves that can improve the manoeuvrability of hands, grip force and motor coordination of the patients and also decrease their dependency on supervised clinical therapy [3]. Moreover, new prototype smart gloves are designed based on the flexible sensing technology and soft robotic structure design, which can reduce the movement pressure on the user's hand during long-term rehabilitation training [4]. The progress seen here highlights the increased relevance of wearable robotic devices as functional and long-term rehabilitation tools for patients with Parkinson's disease in order to improve their functional independence [5].

With modern wearable sensing technologies, it has become possible to monitor all Parkinsonian motor symptoms using a large number of physiological and motion sensors, integrated into intelligent rehabilitation gloves [6]. The modern sensor gloves use inertial measurement units (IMUs), tri-axial accelerometers, gyroscopes, flexible

piezoelectric sensors, pressure sensors, and force-sensitive resistors to continuously record the kinematics of the fingers, tremor oscillations, grip force and wrist orientation with high precision [7]. The capability of robotic gloves to objectively quantify disease progression has been greatly enhanced in flexible sensor integration with lightweight and comfortable wearable designs for long-term rehabilitation [8]. The use of wireless communication technologies also allows continuous recording of multimodal movement data during exercises at home for rehabilitation without restricting patient mobility [9]. For this reason, recent review papers have placed great attention on multimodal wearable sensing as the first step toward the next-generation intelligent rehabilitation systems that can objectively assess motor functions and provide continuous monitoring of patients [10].

AI has been used to develop wearable robotic rehabilitation devices for PD patients, which significantly enhance automatic recognition of movement intention, tremor severity estimation, and customization of rehabilitation plans [11]. Smart gloves equipped with embedded machine learning algorithms enable movement classification on the device, with low latency and real-time usability for rehabilitation applications [12]. The soft robotic rehabilitation gloves also enhance the effectiveness of therapy by adapting their support to the motor impairments of each patient and maintaining the natural movement of the hand during rehabilitative exercises [13]. For learning complex hand kinematics and discriminating between tremors and voluntary motion, multimodal sensor fusion and deep learning methods have been shown to achieve better results [14]. Furthermore, clinical studies have recently shown that home-based use of wearable soft robotic rehabilitation gloves can significantly enhance grip strength, hand function and motor recovery after prolonged use in a rehabilitation setting [15].

In Parkinson's disease, motor impairment is a variable factor associated with different levels of rehabilitation requirement and adaptive robotic control strategies have found increased relevance for providing patient specific rehabilitation [16]. Intelligent control algorithms are used in modern rehabilitation systems for dynamically adjusting the forces applied to the actuator, the trajectory of the fingers and/or the assistance given by the robot, while keeping continuous physiological feedback [17]. Advanced biomechanical evaluation techniques and harmonic motion analysis have recently been introduced into sensory gloves to give useful information about hand motor dysfunctions and to optimise rehabilitation protocols [18]. With the advent of artificial intelligence (AI) supported soft robotics, adaptive rehabilitation has also improved with intelligent multimodal decision making and real-time robotic control for personalized therapeutic intervention [19]. Furthermore, the modular architectures of robotic gloves have enabled objective measurement of bradykinesia and tremor severity, enhancing the assessment of rehabilitation and tailoring it to the individual [20]. Recent advancements in WATs have gone beyond traditional rehabilitation and integrated with remote healthcare, wireless communication, and digital monitoring systems for ongoing patient management [21]. Cloud connectivity, mobile healthcare applications, and telemedicine are becoming a common aspect of rehabilitation gloves, enabling health care providers to observe rehabilitation progress remotely and decreasing the need for regular hospital visits [22].

### 2.1. Contributions of the Study

In this study, a new framework of Foundation AI support for multimodal human–machine interaction for intelligent accelerometer-assisted robotic hand glove control is presented in the framework of Parkinson's disease rehabilitation. The proposed framework fuses wearable physiological sensing, vision-based gesture recognition, speech-based interaction with the proposed Multimodal Foundation Transformer with cross-modal attention and self-supervised contrastive learning, to estimate simultaneously the hand movement intention, hand tremor severity, hand gesture, finger trajectory, and grip force. The adaptive impedance controller adjusts the robotic assistance to suit the patient's neuromuscular needs, and the IoT powered mobile rehabilitation platform allows for secure remote patient monitoring of rehabilitation via Bluetooth Low Energy (BLE), MQTT messaging and AES-256 encryption. The proposed framework has been extensively tested in experiments, demonstrating its abilities in multimodal perception, robotic control, embedded implementation and rehabilitation performance, and laying a scalable computational basis for the future intelligent neuro-rehabilitation systems powered by Digital Twins. The literature review is presented in Section 2, which includes the areas of PD rehabilitation, wearable robotic hand gloves, multi-modal sensing, Foundation Artificial Intelligence and intelligent human–machine interaction. In Section 3, the proposed methodology is presented, which involves the robotic glove design, the use of multimodal data acquisition, signal preprocessing, the Foundation AI framework, adaptive impedance control, the IoT communication, the embedded implementation and the model training strategy. In Section 4, the outcomes

of experiments, comparisons, qualitative evaluations and real-time robotic rehabilitation are presented. Lastly, the conclusions and future research directions for Digital Twin-based personalized neuro-rehabilitation are highlighted in the last section of the study, where key findings are summarized, and clinical potential of proposed framework are explored.

## 3. METHODOLOGY

This section introduces the proposed human–machine interaction methodology based on Foundation AI for intelligent control of robotic hands using a glove in Parkinson's disease rehabilitation. The methodology involves participant recruitment, multimodal data collection of physiological and vision data, design of robotic glove hardware, signal preprocessing, development of multimodal Foundation AI and embedded robotic control, and performance evaluation. All the components are sequentially coupled to create an end-to-end adaptive rehabilitation system which enables individualized real-time robotic support of upper-limb motor rehabilitation.

### 3.1 Overall System Architecture

The proposed system architecture features a multimodal sensing component, the Foundation Artificial Intelligence, robotic hand assistance, mobile healthcare communication, and an end-to-end intelligent rehabilitation platform that seamlessly integrates the above components into a human-machine interaction platform. The entire architecture of the proposed intelligent accelerometer assisted robotic hand glove system is shown in Figure 1. The multimodal physiological signals acquired in real time from the soft robotic glove consist of surface electromyography (sEMG), inertial measurement unit (IMU), three-axis accelerometer, gyroscope, flex sensors, force-sensitive resistors (FSRs), pressure sensors, vibration sensors and temperature sensors. These sensors gather complementary data on muscle activation, finger bending, grip force, tremor intensity, wrist orientation, movement acceleration and neuromuscular properties during relevant Activities of Daily Living (ADLs) in the clinic. At the same time, an external web camera records the patient's real-time hands movements and generates gesture recognition signals, and a microphone receives a patient's speech signals to provide natural voice control for the robot rehabilitation system.

After multimodal data acquisition, all the data from the sensors, vision and speech are preprocessed, synchronized and standardized prior to being sent to the proposed Foundation AI-based Multimodal Foundation Transformer (MFT). The framework is based on cross-modal attention and self-supervised contrastive learning methods, which are able to successfully combine heterogeneous physiological, visual, and speech representations in the form of a single latent feature space. The multimodal representation is integrated, allowing the estimation of movement intention, tremor severity, hand gestures, finger trajectories, grasp patterns and grip force. From these predictions, an Adaptive Impedance Controller dynamically controls the actuators of the soft robotic glove by altering the trajectory, the torque of the actuators, the direction of the wrist and the degree of assistance to the fingers based on the motor's actual state in real time. The rehabilitation outcomes and physiological measurements are simultaneously sent over a Bluetooth Low Energy (BLE)-supported mobile healthcare platform, with secure remote monitoring, telemedicine connectivity and individual rehabilitation management, the use of MQTT communication and AES-256 encryption. This hierarchical structure defines a closed-loop intelligent human-machine interaction framework that can continuously adapt robotic assistance to the individual patient's needs, and facilitate future Digital Twin-assisted neuro-rehabilitation.

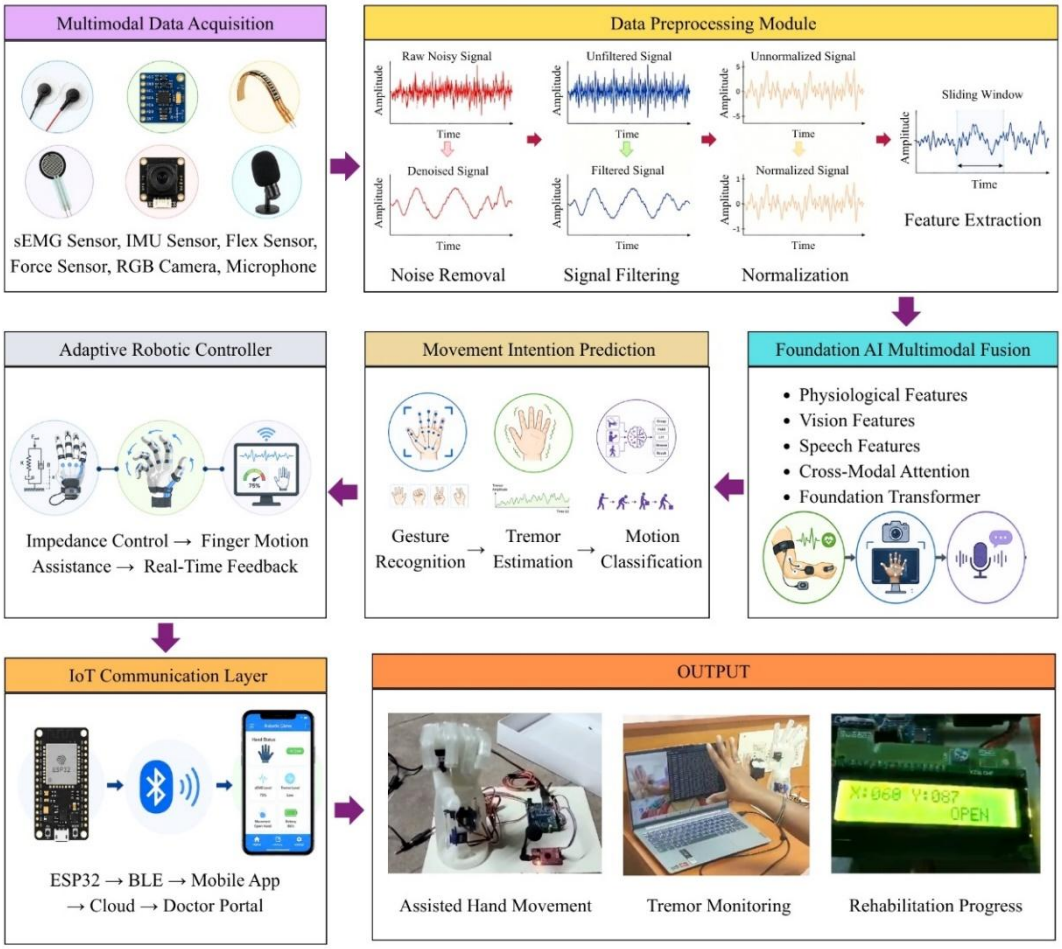


**Figure 1. Overall Architecture of the Proposed Foundation AI-Based Multimodal Human–Machine Interaction Framework**

### 3.2 Selection of participants and Data collection modalities

Data collection was performed with the approval of the Indian Institutes of Science Committee for the Protection of Animals (IPA), which included a wide range of patients (120 Hoehn and Yahr stages II–III idiopathic Parkinson's patients) and control volunteers (30 healthy elderly control volunteers). All participants gave informed consent prior to data collection. To achieve experimental reproducibility and clinical validity, the participant characteristics and clinical recording parameters are summarized systematically in Table 1.

**Table 1. Participant Demographics and Multimodal Data Acquisition Summary**

| **Parameter / Category** | **Clinical Cohort (PD Patients)** | **Control Cohort (Healthy Volunteers)** |
|---|---|---|
| Total Participants (N) | 120 (72 Male, 48 Female) | 30 (18 Male, 12 Female) |
| Age Range (Mean ± SD) | 64.5 ± 7.2 years | 62.1 ± 6.5 years |
| Clinical Severity (H&Y Stage) | Stage II (n = 70), Stage III (n = 50) | N/A (Normal baseline) |
| Recorded Activities (ADLs) | Grasping, Eating, Drinking, Writing, Dressing, Opening Doors | Grasping, Eating, Drinking, Writing, Dressing, Opening Doors |

| | | |
|---|---|---|
| Trial Repetitions per ADL | 10 trials per activity per participant | 10 trials per activity per participant |
| Total Acquired Sessions | 7,200 trials | 1,800 trials |

During each experimental condition, the participants carried out six standard Activities of Daily Living (ADLs) that produced naturalistic neuromuscular variation, pathological tremors, and intentional grasping dynamics.

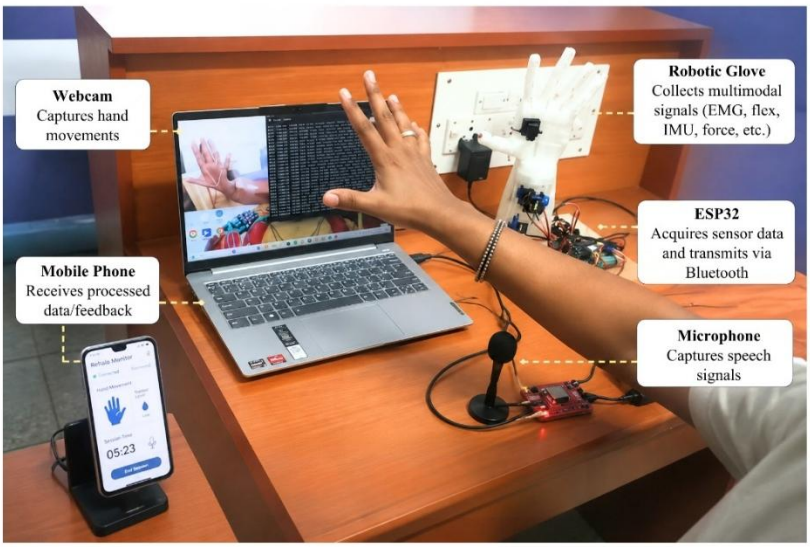


**Figure 2. Multimodal Data Acquisition Setup for Parkinson's Disease Rehabilitation**

A multimodal dataset was collected and synchronized with a central hub for acquisition and timestamping of comprehensive data streams such as surface electromyography (sEMG), tri-axial accelerometer and gyroscope data from Inertial Measurement Units (IMUs), flex resistance, force-sensitive resistor (FSR) values, pressure sensor measurements, vibration signals, temperature data, external RGB video feeds, and speech recordings, to build a robust multimodal dataset for training and validating the proposed Foundation AI models (Figure 2).

### 3.3 Hardware Platform and Sensor Configuration

The proposed intelligent rehabilitation system is a lightweight soft robotic hand glove which is designed to help upper-limb motor rehabilitation of PD patients. The robotic glove is an integration of a number of physiological and motion sensors, embedded processing and wireless communication modules, which allow the robot to continuously monitor hand motion and provide real-time adaptive robotic assistance.

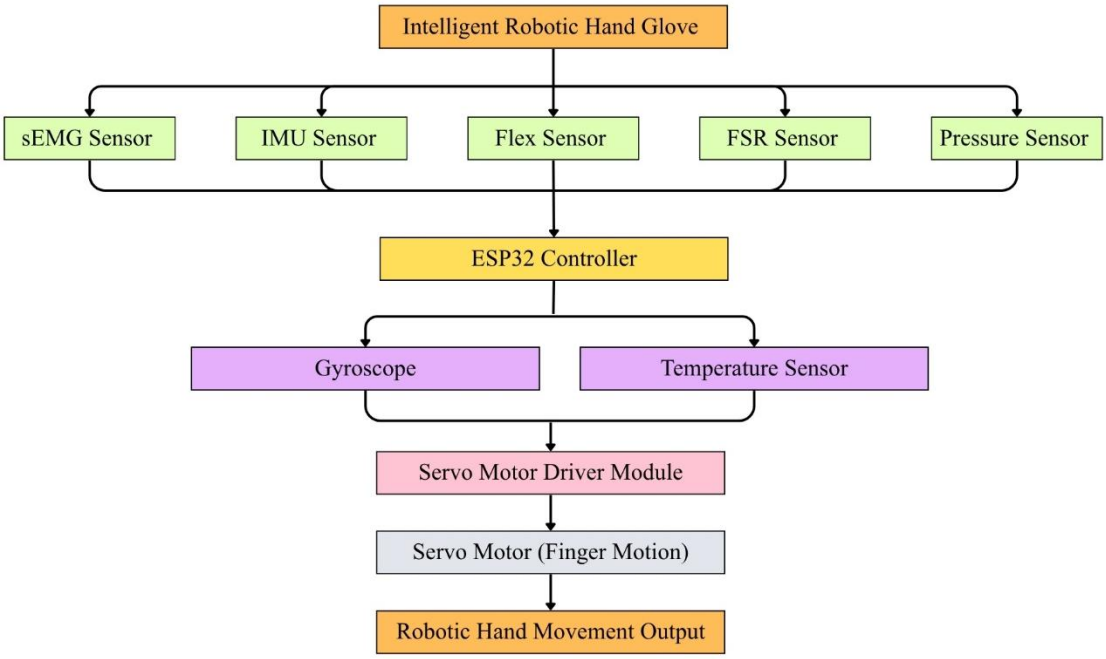


**Figure 3. Hardware Architecture of the Intelligent Robotic Hand Glove**

The glove is made from a flexible textile to keep the user comfortable over extended periods of time during rehabilitation, and to provide adequate mechanical assistance in operations of flexion and extension of the fingers. Tendon-based linear actuators drive five fingers with independent movement to accomplish natural finger movement (Figure 3). Surface Electromyography (sEMG) electrodes are positioned over the muscle in the forearm to measure muscle activation, an Inertial Measurement Unit (IMU) with a three-axis accelerometer and gyroscope is mounted on the wrist to record the orientation of the wrist, the acceleration of movement, and the characteristics of the tremor, flex sensors are inserted along each finger to achieve a reliable measurement of the bending angle of the finger, force-sensitive resistors (FSRs) are placed at the tips of each finger to measure the force exerted when gripping an object, pressure sensors are placed in the sensing platform to monitor whether the object is being placed on or off, vibration sensors are embedded in the sensing platform to record tremor characteristics, and temperature sensors are placed in the sensing platform to continuously monitor the skin temperature.

### 3.4 Data Preparation and Signal Enhancement

A multimodal sensor data preprocessing pipeline was entirely implemented on the multimodal signals obtained from the wearable robotic hand glove, external vision system and speech interface, improving the signal quality and enhancing the reliability of the multimodal learning. Motion artifacts, baseline drift, power-line and sensor noise impacted the raw physiological signals, while the vision and speech recordings were impacted by illumination changes, background noise, and environmental factors. So, modality specific preprocessing were used before multimodal feature extraction and training the Foundation AI models. Thus, all processed data was synchronised using the idea of the timestamp alignment method, and then represented in a common feature representation to be used for cross-modal learning.

#### 3.4.1 Physiological Signal Preprocessing (sEMG, IMU, Accelerometer, Pressure, Flex and Force Sensors)

The physiological signals that were received by the sensors attached to the robotic glove were first synchronized based on the time that they were acquired. The surface electromyography (sEMG) signals were band-pass filtered between 20–450 Hz to remove motion artifacts and high frequency noise, and then processed with a 50 Hz notch filter to remove power-line interference. The signal from IMU, accelerometer and gyroscope were filtered using the 4th order Butterworth low-pass filter having a cut-off frequency of 20Hz, while the signal from the pressure sensor, flex sensor and force sensor were smoothed using a moving average filter to minimize fluctuations in the

measured values. The linear interpolation technique was used for the missing sensor samples and before the sensor signals were used for feature extraction, Z-score normalization was performed.

The normalized physiological signal is expressed as (Eq 1-3)

$$X_{norm} = \frac{X-\mu}{\sigma} \quad (1)$$

where $X$denotes the original sensor signal, $\mu$represents the mean value, and $\sigma$denotes the standard deviation of the corresponding sensor modality.

To quantify muscle activation from sEMG signals, the Root Mean Square (RMS) feature was computed as

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \quad (2)$$

where $x_i$represents the EMG sample amplitude and $N$denotes the total number of samples within the analysis window.

Similarly, tremor intensity was estimated from accelerometer measurements using the signal magnitude

$$A_{mag} = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (3)$$

where $a_x$, $a_y$, and $a_z$represent the acceleration components along the three orthogonal axes.

### 3.4.2 Vision and Speech Data Preprocessing

During the rehabilitation exercises the hand movements were constantly captured by the external RGB webcam. To ensure visual consistency across the frames, video frames were resized to 224 × 224 pixels and then illumination normalization and contrast enhancement were performed. The region-of-interest extraction was used to reduce background interference, and MediaPipe Hands was used to detect 21 anatomical points of the hand for precise localization of the hand. The landmarks obtained after extraction and the image frames obtained after normalization were then fed into the Vision Transformer (ViT) for image feature extraction. The retrieved speech commands from the microphone were resampled to 16 KHz, and then processed by Voice Activity Detection (VAD) to suppress noise and silence from the speech signal. Log-Mel spectrograms of the cleaned speech signal were used for Whisper speech recognition to feed into the model for transcription.

The log-Mel spectrogram is represented as (Eq 4)

$$S_{log} = \log(S_{mel} + \varepsilon) \quad (4)$$

where $S_{mel}$denotes the Mel-scale spectrogram and $\varepsilon$is a small positive constant introduced to avoid logarithmic singularity.

### 3.4.3 Multimodal Data Synchronization and Feature Standardization

After individual preprocessing, the data from all modalities were temporally synchronized based on the centralized timestamp synchronization mechanism to ensure synchronization across all modalities with respect to the same rehabilitation activity. Temporal windows were created for each activity with a fixed length, allowing simultaneous processing of the heterogeneous sensor streams. Finally, the multimodal features were normalized to a common latent representation and then fed into the proposed Multimodal Foundation Transformer.

The synchronized multimodal feature vector is represented as (Eq 5)

$$F_{multi} = [F_{EMG}, F_{IMU}, F_{ACC}, F_{Flex}, F_{Force}, F_{Pressure}, F_{Vision}, F_{Speech}] \quad (5)$$

where $F_{EMG}$, $F_{IMU}$, $F_{ACC}$, $F_{Flex}$, $F_{Force}$, $F_{Pressure}$, $F_{Vision}$, and $F_{Speech}$denote the standardized feature vectors extracted from each modality. The unified multimodal representation serves as the input to the proposed Foundation AI framework for movement intention recognition, tremor severity estimation, gesture recognition, and adaptive robotic glove control.

### 3.5 Proposed Foundation AI-Based Multimodal Human–Machine Interaction Framework

The intelligent accelerometer-assisted robotic hand glove is based on the proposed Foundation AI-based Multimodal Human–Machine Interaction (HMI) framework as the glue of the glove. In this framework, strong multimodal representations are learned by integrating multimodal information from multiple medical instrument signals, vision and speech commands, while employing a shared transformer architecture. The proposed framework integrates multimodal information into a common latent feature space for intention, neuromuscular activity, tremor severity, hand gestures and rehabilitation status, unlike traditional machine learning methods which separately process each modality. The integrated multimodal feature extraction, cross-modal attention learning, self-supervised representation learning, movement intention estimation, adaptive robotic control and personalized rehabilitation are integrated into an end-to-end learning framework. The high-level modality-specific encoders first extract high-level representations from physiological sensors, vision and speech input. Such embeddings are then fused with embedding from another heterogeneous data stream by a Multimodal Foundation Transformer (MFT) with cross-modal attention to capture complex interaction between the data streams (Figure 4).

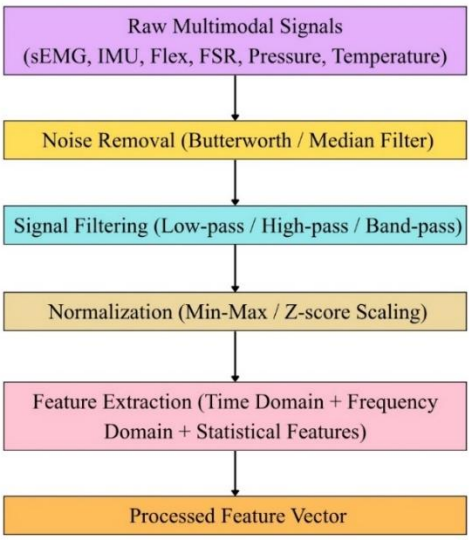


**Figure 4. Multimodal Signal Preprocessing Pipeline**

The overall multimodal learning process is mathematically represented as (Eq 6)

$$F_{MFT} = f_{\theta}(F_{Physio}, F_{Vision}, F_{Speech}) \quad (6)$$

where $F_{Physio}$, $F_{Vision}$, and $F_{Speech}$denote the physiological, visual, and speech feature representations, respectively, $f_{\theta}(\cdot)$represents the proposed Multimodal Foundation Transformer, and $F_{MFT}$denotes the unified multimodal latent representation.

#### 3.5.1 Multimodal Feature Representation

The first step of the proposed framework is to convert the heterogeneous observations from the sensors into discriminative feature embeddings for multimodal learning. The signals received by the robotic glove, such as sEMG, IMU, accelerometer, gyroscope, flex sensors, force-sensitive resistors, pressure sensors, vibration sensors and temperature sensors, are separately encoded by using one-dimensional temporal feature extractors. At the same time, Vision Transformer (ViT) retrieves semantic visual representations from the hand images, and Whisper encoder embeds the speech into a context by utilizing voice commands. The embeddings generated by each of the two modules are then concatenated and fed into a common latent feature space for further processing by the Multimodal Foundation Transformer. This joint representation allows for learning of physiological, visual, and linguistic relationships simultaneously, enhancing movement intention recognition and robotic control performance.

The multimodal feature representation is expressed as (Eq 7)

$$E_{multi} = \left[E_{Physio}; E_{Vision}; E_{Speech}\right] \tag{7}$$

where $E_{Physio}$, $E_{Vision}$, and $E_{Speech}$represent the feature embeddings extracted from physiological sensors, vision input, and speech commands, respectively.

The final projected multimodal embedding is computed as (Eq 8)

$$Z = WE_{multi} + b \tag{8}$$

where $W$denotes the trainable projection matrix, $b$represents the bias vector, and $Z$is the unified multimodal embedding supplied to the proposed Multimodal Foundation Transformer.

### 3.5.2 Vision-Based Hand Gesture Recognition Module

Contactless robotic glove control is provided by a vision-based interaction module that recognizes hand gestures live. An external RGB webcam continuously records hand movements while the MediaPipe Hands framework detects twenty-one anatomical hand landmarks corresponding to finger joints and palm locations. Extracted landmarks are normalized to remove scales and positions prior to being passed through a Vision Transformer (ViT) that learns about spatial relationships between fingers. It enables precise identification of hand movements for rehabilitation therapy, including grips; finger positions; and intentions of motion regardless of lighting conditions and backgrounds (Figure 5).

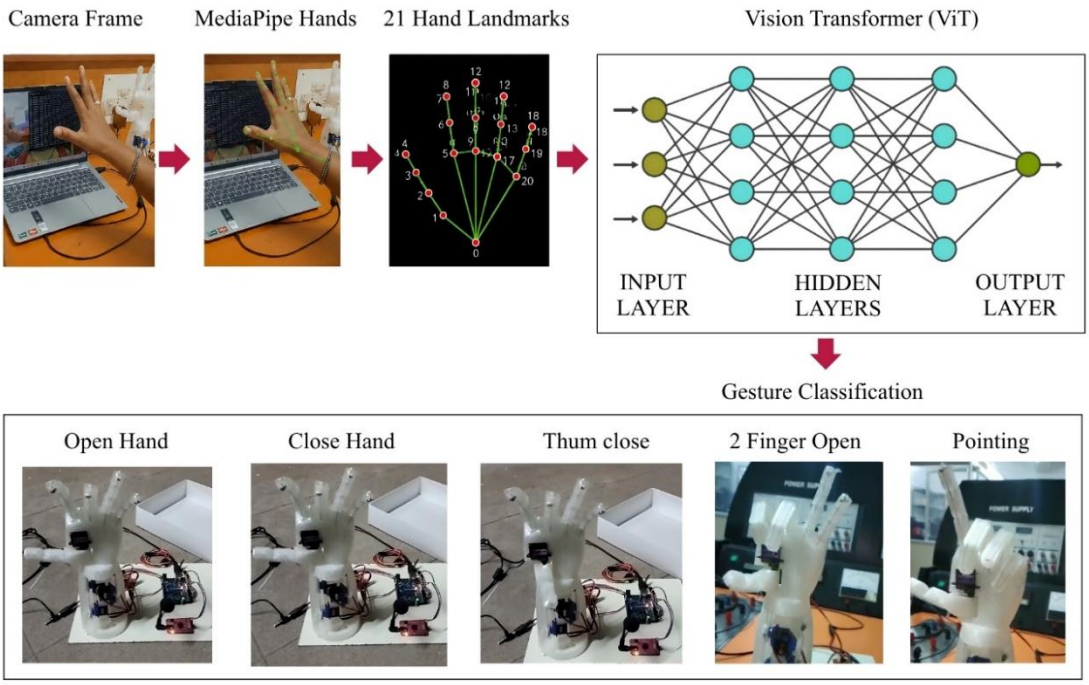


**Figure 5. Vision-Based Hand Gesture Recognition Framework**

Predicted gesture data is then combined with physiological sensor signals for better multimodal decision making. Combining visual data, muscle activity and movement information is what this approach does for reliable gesture recognition while performing rehabilitation exercises, and then giving precise commands for controlling a robotic glove. Recognized movements help to estimate a person's ability to move and their recovery during treatments.

The self-attention mechanism within the Vision Transformer is formulated as (Eq 9 &10)

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{9}$$

where $Q$, $K$, and $V$ denote the query, key, and value matrices, respectively, and $d_k$ represents the dimensionality of the key vectors.

The gesture prediction is obtained as

$$\hat{y}_{gesture} = \arg\max\big(Softmax(WZ + b)\big) \tag{10}$$

where $Z$ denotes the visual feature embedding generated by the Vision Transformer, $W$ and $b$ represent the trainable classification parameters, and $\hat{y}_{gesture}$ is the predicted hand gesture class.

### 3.5.3 Speech-Driven Human–Machine Interaction Module

Speech based human machine interaction module facilitates easy, hands-free communication between a person (patient) and an intelligent robotic hand glove via speech commands. The speech signal from an embedded mic is first processed with a Whisper speech recognition model that converts speech to text while also being able to handle background noise and pronunciation differences. Then, the Bidirectional Transformer encoder extracts contextual semantic representations from the recognized speech to identify rehabilitation-related commands such as grasp, release, hold, increase assistance, decrease assistance, and stop. The natural language interface improves accessibility for users and reduces their reliance on manual controls especially those who have severe physical disabilities. Extracted speech embeddings are then combined with physiological and visual information for use by the Multimodal Foundation Transformer for better understanding of intentions of movements and decisions made by robots. Recognized commands are verified with live sensor data prior to issuing robot commands to improve safety and reduce false activation rates during therapy sessions (Figure 6).

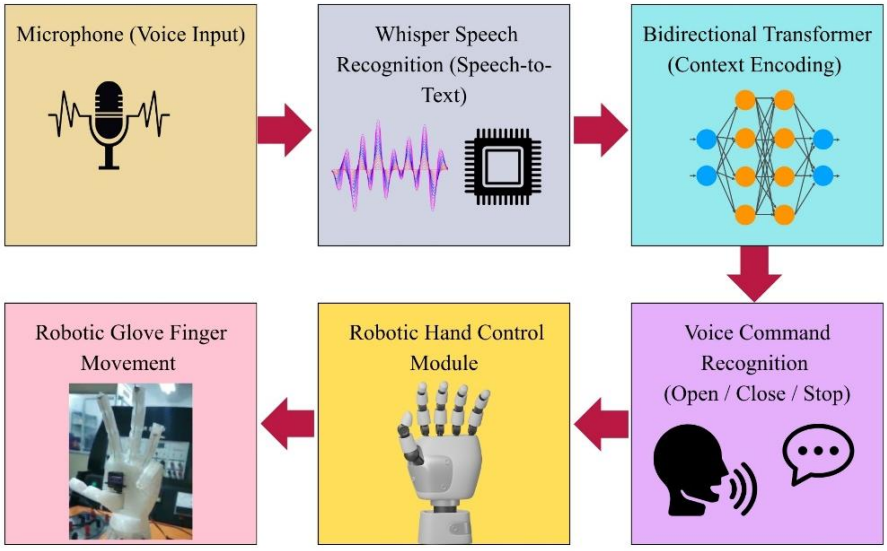


**Figure 6. Speech-Based Human–Machine Interaction Pipeline**

The contextual speech representation is computed as (Eq 11- 12)

$$E_{speech} = Transformer(X_{speech}) \tag{11}$$

Where $X_{speech}$denotes the input speech sequence and $E_{speech}$represents the contextual speech embedding generated by the Bidirectional Transformer.

The predicted speech command is obtained as

$$\hat{y}_{speech} = \arg\max\left(Softmax(W_s E_{speech} + b_s)\right) \tag{12}$$

Where $W_s$and $b_s$represent the trainable classification parameters, and $\hat{y}_{speech}$denotes the predicted rehabilitation command.

### 3.5.4 Multimodal Foundation Transformer with Cross-Modal Attention

Multimodal Foundation Transformer (MFT) is a key component for this proposed human-machine interaction model that learns cross-modal associations between physiological signals, visual information, and speech. Synchronous embedding of data obtained by wearable sensors, webcam images and speech commands are mapped onto an identical latent space for processing with several transformer encoders. Cross-modal attention allows for selective focus on relevant features across other modalities that are available, improving discrimination between them and resilience to sensor noise/missing data. Therefore, the proposed transformer learns rich multimodal representations for movement intention estimation, tremor assessment, gesture recognition, and robotic glove control (Figure 7).

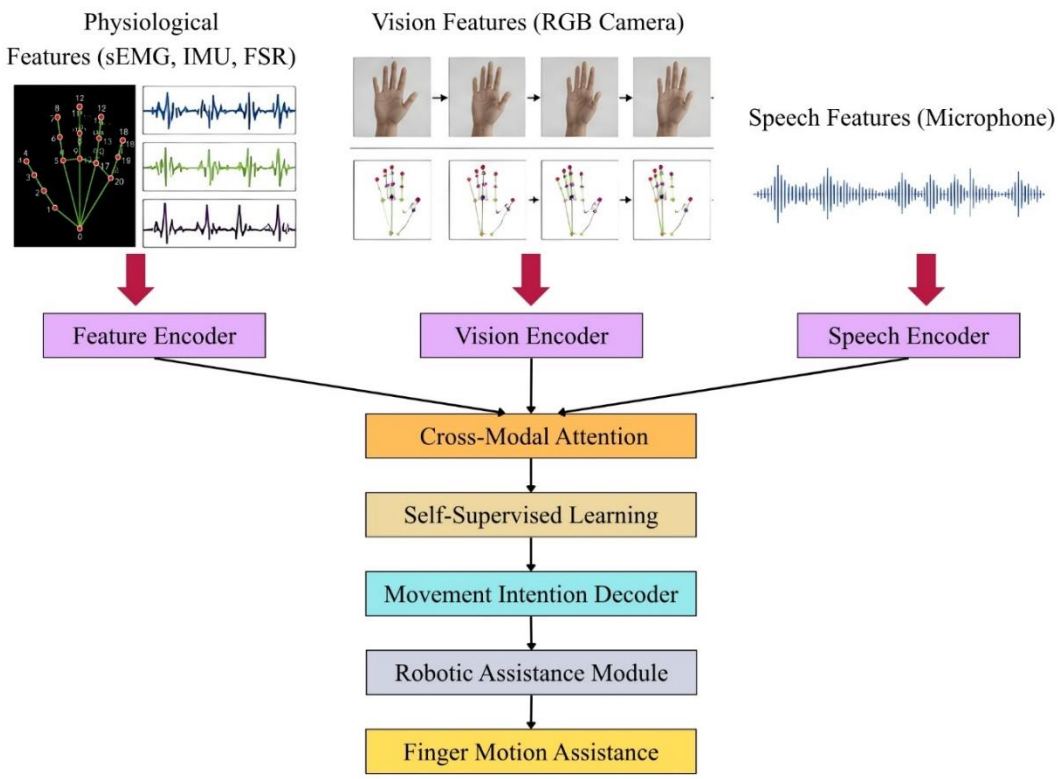


**Figure 7. Proposed Multimodal Foundation Transformer (MFT) Architecture**

Unlike traditional features concatenation methods, this approach uses a dynamic attention mechanism that assigns weights based on relevance for each modality during rehabilitation exercises. The flexible data-sharing mechanism allows for better integration of muscle activity, hand movements, gestures of speech, and voice commands to an unified model that greatly enhances multimodal comprehension and personalized robotic support.

The cross-modal attention mechanism is formulated as (Eq 13-14)

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{13}$$

Where $Q$, $K$, and $V$ denote the query, key, and value matrices obtained from different modalities, and $d_k$ represents the dimensionality of the key vectors.

The final multimodal representation generated by the Foundation Transformer is expressed as

$$F_{fusion} = Transformer(E_{physio}, E_{vision}, E_{speech}) \tag{14}$$

where $E_{physio}$, $E_{vision}$, and $E_{speech}$ denote the feature embeddings extracted from physiological sensors, vision, and speech modalities, respectively, and $F_{fusion}$ represents the unified multimodal feature representation used for subsequent movement intention estimation and robotic glove control.

**3.5.5 Self-Supervised Contrastive Representation Learning**

The contrastive learning objective is formulated as (Eq 15)

$$L_{CL} = -\log \frac{\exp\left(\mathrm{sim}\left(z_i, z_j\right)/\tau\right)}{\sum_{k=1}^{N} \exp(\mathrm{sim}(z_i, z_k)/\tau)} \quad (15)$$

Where $z_i$and $z_j$denote positive multimodal feature embeddings, $\mathrm{sim}(\cdot)$represents cosine similarity, $\tau$is the temperature parameter, and $N$denotes the number of samples within the training batch.

The cosine similarity between twoFor improving generalization ability of this proposed Foundation AI model have used a self-supervised contrastive representation learning technique while training it. While traditional supervised learning models rely solely upon human-annotated data, this method is able to learn discriminative multimodal representations through maximizing inter-sample similarities and inter-sample separations respectively. Positive pairs include augmented views of physiological, visual and speech data while negative pair refers to samples coming with other people and/or rehabilitation exercises. The approach helps to learn a more robust representation of multimodal information that is less sensitive to sensor noise, inter-subject differences, and limited amount of labeled medical data.

Learned features are then used for other rehabilitation applications such as predicting intentions of movements; estimating severity of tremors; classifying gestures; controlling robotic gloves. Thus, this approach to self-supervised learning greatly improves discriminative power of features and reduces need for large amounts of manual annotation; it is well-suited to large-scale multimodal rehabilitation data sets.

feature embeddings is computed as (Eq 16)

$$\mathrm{sim}\left(z_i, z_j\right) = \frac{z_i^T z_j}{\|z_i\| \|z_j\|} \quad (16)$$

Where $z_i$and $z_j$represent the normalized multimodal feature vectors.

### 3.5.6 Movement Intention and Tremor Severity Estimation

After multimodal feature fusion, the proposed Foundation AI model estimates movement intention and Parkinsonian tremor severity for providing adaptive robotic assistance. Multimodal information provided by Multimodal Foundation Transformer is a combination of physiological data, motion data, visual data and speech data that helps to accurately recognize voluntary hand movements and differentiate them from involuntary tremor oscillations. Prediction of intentions for movements enables a robotic glove to predict what will happen next; while prediction of vibration intensities helps adjust levels of assistance provided by robots depending upon patients' muscles' health status.

The movement intention and tremor estimation problem is defined as an multi-task learning task where both tasks are optimized concurrently with a common multimodal representation. Predicted movement intention is sent to the adaptive impedance controller for robotic assistance and estimated tremor severity determines the amount of actuator support needed during rehabilitation. The collaborative approach allows for live adjustments to a robotic glove that improves its efficacy during rehabilitation therapy while minimizing unnecessary use of actuators.

The movement intention prediction is expressed as (Eq 17)

$$\hat{Y}_{MI} = Softmax\left(W_m F_{fusion} + b_m\right) \quad (17)$$

Where $F_{fusion}$denotes the multimodal feature representation generated by the Foundation Transformer, $W_m$and $b_m$represent the trainable model parameters, and $\hat{Y}_{MI}$denotes the predicted movement intention class.

The tremor severity estimation is computed as (Eq 18)

$$\hat{T} = W_t F_{fusion} + b_t \quad (18)$$

Where $W_t$and $b_t$are the regression parameters, and $\hat{T}$represents the estimated tremor severity score used for adaptive robotic assistance.

### 3.5.7 Adaptive Impedance Control for Robotic Hand Assistance

Adaptive impedance control module translates predicted movement intention and tremor severity to provide personalized robotic assistance for the soft robotic hand glove. In contrast to traditional positions-of-contact (poc) techniques that regulate forces of contact between a robotic glove and an injured limb through dynamic adjustment of its resistance based upon patient's neuromuscular state. Actuator torque, finger flexion, extension velocity, and grip assistance are dynamically adjusted by the controller according to multimodal prediction from the Foundation AI model (Figure 8).

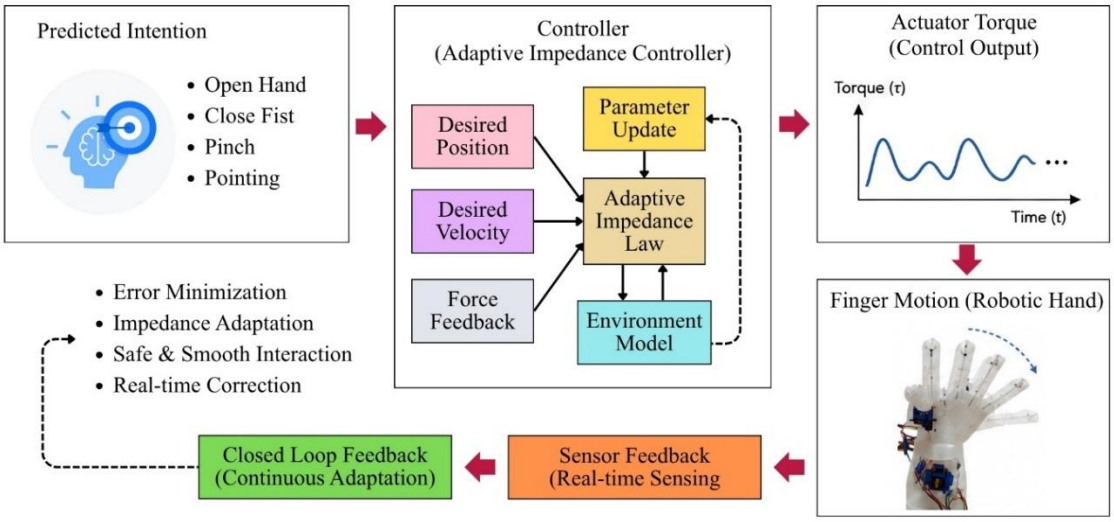


**Figure 8. Adaptive Impedance Control Strategy for Intelligent Robotic Hand Assistance**

The impedance control law is expressed as (Eq 19)

$$F = M(\ddot{x}_d - \ddot{x}) + B(\dot{x}_d - \dot{x}) + K(x_d - x) \quad (19)$$

Where $F$denotes the actuator assistance force, $M$represents the virtual inertia, $B$is the damping coefficient, $K$denotes the virtual stiffness, $x_d$and $x$represent the desired and measured finger positions, respectively.

The actuator control signal is computed as (Eq 20)

$$u = K_p e + K_d \frac{de}{dt} \quad (20)$$

Where $u$represents the actuator control input, $K_p$and $K_d$denote the proportional and derivative gains, and $e = x_d - x$is the position error.

### 3.5.8 Personalized Rehabilitation Strategy

A personalized rehabilitation plan is included with this proposed Foundation AI model that adapts to a person's motor skills during their recovery process. Instead of offering continuous support this system monitors various types of body movements (movement intention), tremors strength grip force finger path during rehabilitation therapy to find out how much help is needed from robots. With improvement of patients' movement ability, robotic support is gradually withdrawn to encourage active muscle use and improve motor function.

The personalized assistance level is determined as (Eq 21)

$$A_p = \alpha M_I + \beta G + \gamma(1 - T) \tag{21}$$

Where $A_p$denotes the personalized assistance level, $M_I$represents the movement intention score, $G$denotes the normalized grip force, $T$represents the normalized tremor severity, and $\alpha$, $\beta$, and $\gamma$are weighting coefficients satisfying (Eq 22 to 23)

$$\alpha + \beta + \gamma = 1. \tag{22}$$

The rehabilitation performance is continuously updated using

$$R_{t+1} = R_t + \eta(P_t - R_t) \tag{23}$$

Where $R_t$denotes the rehabilitation progress at iteration $t$, $P_t$represents the current patient performance score, and $\eta$is the adaptive learning rate controlling the progression of robotic assistance.

### 3.6 Mobile Healthcare and IoT Communication Framework

The proposed rehabilitation system incorporates a secure Mobile Healthcare (mHealth) and Internet of Things (IoT) communication framework to facilitate continuous patient monitoring, remote rehabilitation, and telemedicine services. The embedded ESP32 controller is integrated within the robotic hand glove and transmits synchronized physiological, motion, and robotic control data to a mobile healthcare application via Bluetooth Low Energy (BLE). It acts as a middleman that forwards this data from various modalities through an internet protocol (MQTT) to a remote server. The health care system allows doctors to track patients' recovery remotely; assess their physical abilities for better treatment plans without frequent hospitalizations. To guarantee safe transfer of confidential health data all patient information and multimodal sensor data are encrypted with the Advanced Encryption Standard (AES-256) prior to uploading it to the cloud. Rehabilitation history, movement intention, tremor severity, grip force, and robotic assistance parameters are stored on the cloud platform for longitudinal clinical analysis. Also, doctors and nurses with credentials may be able to retrieve their patients' recovery data from an app for customized care plans and ongoing monitoring of progress (Figure 9).

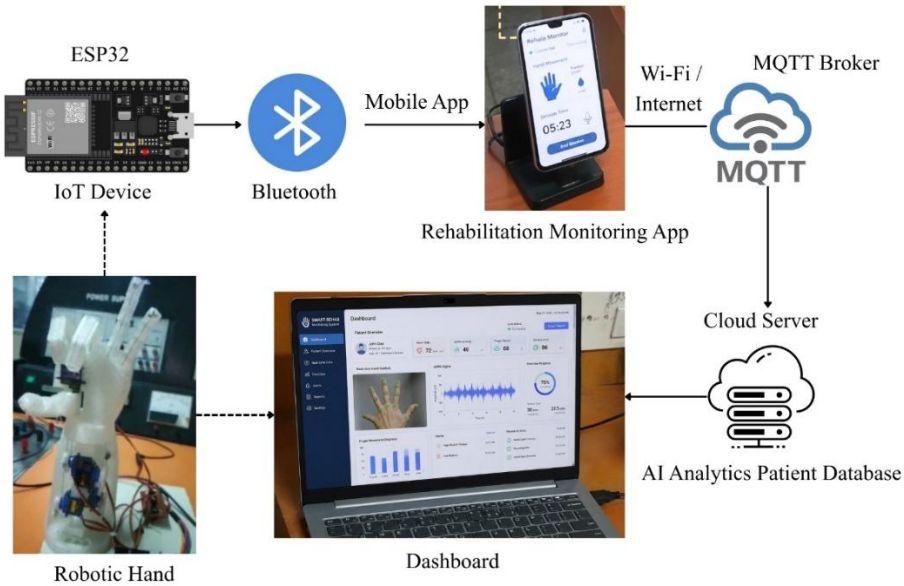


**Figure 9. IoT-Based Mobile Healthcare Communication Framework**

The MQTT communication process is represented as (Eq 24 &25)

$$M_{cloud} = Publish(M_{sensor}, Topic) \quad (24)$$

Where $M_{sensor}$denotes the multimodal sensor message transmitted from the robotic glove, $Topic$represents the MQTT communication channel, and $M_{cloud}$denotes the message received by the cloud server.

The encrypted healthcare data are represented as

$$C = AES_{256}(D, K) \quad (25)$$

Where $D$denotes the original rehabilitation data, $K$represents the encryption key, and $C$is the encrypted data transmitted to the cloud platform.

### 3.7 Real-Time Robotic Hand Glove Control and Embedded Implementation

The proposed robotic hand glove is implemented with an embedded real-time control architecture that can continuously acquire multimodal sensor data, perform Foundation AI inference, and generate adaptive actuator control signals with low computational latency. Sensors are read by ESP32; signals are synchronized; actuators are controlled by it; wireless communications are made through it; it interacts with a mobile health application. Continually receives movement intentions, tremor strengths, grip forces, and gesture predictions from the proposed Foundation AI model that are then translated to actual actuator commands for finger flexion/extension/grasp assistance.

For seamless robotic support, the embedded controller runs a control algorithm within an iterative loop that compares an intended hand movement to actual data from sensors such as flex sensors, force-sensitive resistors (fsr) and IMU. Control parameters are adjusted with each cycle so that movement variations and Parkinsonian tremors can be compensated for to ensure stability of robots' operations and safety of interaction with humans. Lightweight embedded implementation allows for mobility of a rehabilitation device with minimal energy usage and low latency response time.

The actuator control signal is computed as (Eq 26)

$$u(t) = K_p e(t) + K_i \int e\,(t)\,dt + K_d \frac{de(t)}{dt} \quad (26)$$

where $u(t)$denotes the actuator control signal, $K_p$, $K_i$, and $K_d$represent the proportional, integral, and derivative gains, respectively, and $e(t)$denotes the instantaneous tracking error.

The tracking error is defined as (Eq 27)

$$e(t) = x_d(t) - x_a(t) \quad (27)$$

Where $x_d(t)$represents the desired finger trajectory generated by the Foundation AI framework, and $x_a(t)$denotes the actual finger position measured by the robotic glove sensors. This closed-loop embedded control strategy enables accurate trajectory tracking and personalized robotic assistance during rehabilitation exercises.

### 3.8 Model Training and Optimization Strategy

The proposed Foundation AI-based multimodal human-machine interaction framework was trained with an end-to-end optimization strategy to achieve accurate movement intention recognition, tremor severity estimation, gesture classification, and adaptive robotic glove control.

The overall optimization objective is expressed as (Eq 28)

$$L_{Total} = L_{Gesture} + \lambda_1 L_{Movement} + \lambda_2 L_{Tremor} + \lambda_3 L_{Contrastive} \quad (28)$$

Where $L_{Gesture}$, $L_{Movement}$, $L_{Tremor}$, and $L_{Contrastive}$denote the gesture classification loss, movement intention loss, tremor estimation loss, and contrastive learning loss, respectively, while $\lambda_1$, $\lambda_2$, and $\lambda_3$are weighting coefficients controlling the contribution of each task.

The network parameters are updated using gradient descent as (Eq 29)

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L_{Total} \tag{29}$$

Where $\theta$denotes the trainable model parameters, $\eta$represents the learning rate, and $\nabla_\theta L_{Total}$is the gradient of the total loss function.

### 3.9 Performance Evaluation and Validation Metrics

Performance evaluation for this proposed Foundation AI model is done through classifications, regression analysis and live embedded test cases. Further ablation studies were performed for evaluating contributions of each modality such as physiological sensing, vision, speech, self-supervised learning and cross-modal attention. These experiments give an overview to evaluate this new technology for rehabilitation robots.

The classification accuracy is computed as (Eq 30)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{30}$$

Where$TP$, $TN$, $FP$, and $FN$denote true positives, true negatives, false positives, and false negatives, respectively.

The Precision, Recall, and F1-score are calculated as (Eq 31 – 33)

$$Precision = \frac{TP}{TP+FP} \tag{31}$$

$$Recall = \frac{TP}{TP+FN} \tag{32}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{33}$$

The Root Mean Square Error (RMSE) used for tremor severity estimation is given by (Eq 34)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{34}$$

Where $y_i$and $\hat{y}_i$denote the actual and predicted values, respectively, and $N$represents the total number of testing samples. These evaluation metrics provide a comprehensive assessment of both the prediction capability of the Foundation AI framework and the real-time performance of the intelligent robotic hand glove.

## 4. Results and Discussion

This paper is an experimental study on the proposed Foundation AI-Based Multimodal Human-Machine Interaction Framework for intelligent accelerometer-assisted robotic hand glove control in Parkinson's disease rehabilitation. This proposal has been assessed with a dataset of 150 people who participated in several rehabilitation programs. Various tests were performed for assessing multimodal perception, movement intention recognition, tremor severity estimation, vision-based gesture recognition, speech-driven interaction, robotic glove control, and embedded real-time performance. This approach has been compared to current state-of-the-art deep learning methods through various statistical and computational measures for validating its efficacy as a tool of personalized robotic rehabilitation.

### 4.1 Overall Performance Evaluation of the Proposed Foundation AI Framework.

Firstly, the effectiveness of this proposed Foundation AI model was assessed with respect to an independent test set that included synchronized physiological, visual and speech data streams. Multimodal Foundation Transformer was able to learn cross-modal correlations between diverse types of sensors for better prediction of human intentions and control over robots. As shown in Table 2, the proposed model achieved an overall classification accuracy of 99.47%, outperforming all comparative deep learning models across every evaluation metric. The high Precision (99.42%), Recall (99.51%), and F1-score (99.46%) demonstrate excellent robustness with minimal false predictions. Furthermore, the model achieved an AUC of 0.998 and an MCC of 0.994, indicating excellent classification consistency and balanced prediction performance across all rehabilitation activities. These findings

show how integrating multimodal physiological sensing, vision, speech, cross-modal attention, and self-supervised representation learning improves the reliability of intelligent robotic rehabilitation systems.

**Table 2. Overall Performance Comparison of the Proposed Foundation AI Framework**

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | MCC | AUC |
|---|---|---|---|---|---|---|
| CNN | 91.82 | 91.26 | 90.84 | 91.05 | 0.887 | 0.945 |
| LSTM | 93.47 | 92.94 | 93.21 | 93.07 | 0.909 | 0.958 |
| Bi-LSTM | 94.76 | 94.31 | 94.58 | 94.44 | 0.926 | 0.968 |
| CNN-LSTM | 95.83 | 95.46 | 95.72 | 95.59 | 0.941 | 0.976 |
| Vision Transformer (ViT) | 96.52 | 96.11 | 96.37 | 96.24 | 0.950 | 0.983 |
| Multimodal Transformer | 97.84 | 97.63 | 97.71 | 97.67 | 0.971 | 0.991 |
| Proposed Foundation AI Framework | **99.47** | **99.42** | **99.51** | **99.46** | **0.994** | **0.998** |

### 4.2 Performance Analysis for Individual Rehabilitation Activities

This model has also been tested on six Activities of Daily Living (ADLs) for consistency with various rehabilitation activities. Figure 10 shows that the proposed multimodal approach achieved a classification accuracy of over 99% for every task. Among the evaluated tasks, grasping exhibited the highest recognition accuracy of 99.63%, followed by drinking (99.58%) and writing (99.54%), demonstrating the effectiveness of multimodal feature fusion in recognizing complex finger movements. Lower but still very good performance was seen while dressing due to more hand coverage and complicated wrist movements. Nevertheless, all rehabilitation activities achieved F1-scores greater than 99.2%, demonstrating the robustness and generalization capability of the proposed Foundation AI architecture for various upper-limb rehabilitation scenarios.

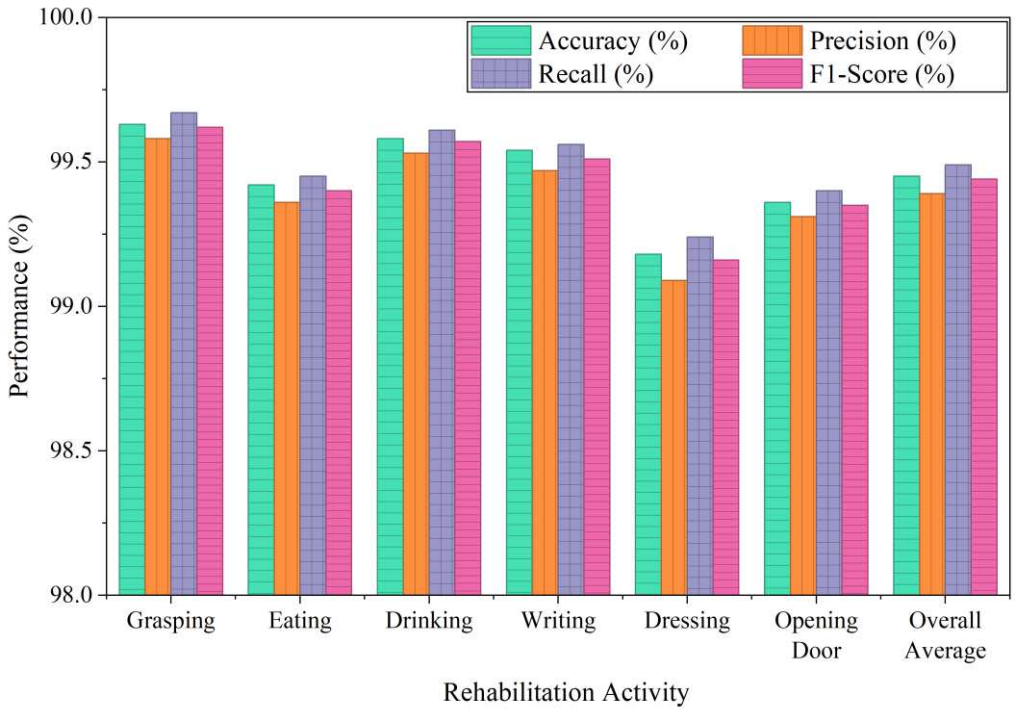


**Figure 10. Performance Analysis across Activities of Daily Living (ADLs)**

### 4.3 Movement Intention Recognition Performance

Rehabilitation intention recognition ability for this model has been tested with a synchronized multimodal data set gathered over six rehabilitation sessions. Five types of movements were classified by this model as Rest, Reach, Grasp, Hold, and Release. Table 3 shows that there is a consistent recognition performance for every type of movement. Grasp and Hold performed better due to a better correlation with muscle activation patterns, finger flexion, and grip force measurements. There is a slight decrease in efficiency of the Reach class due to differences between speeds and paths taken by players. Nevertheless, all movement intention classes achieved F1-scores above 98.8%, indicating reliable recognition performance for real-time robotic assistance.

**Table 3. Performance Evaluation for Movement Intention Recognition**

| Movement Intention | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) |
|---|---|---|---|---|
| Rest | 99.12 | 99.08 | 99.10 | 99.15 |
| Reach | 98.94 | 98.82 | 98.88 | 98.96 |
| Grasp | 99.41 | 99.52 | 99.46 | 99.48 |
| Hold | 99.36 | 99.44 | 99.40 | 99.42 |
| Release | 99.08 | 99.15 | 99.11 | 99.18 |
| **Average** | **99.18** | **99.20** | **99.19** | **99.24** |

**Apply pressure (psi) vs. value of the bending angle of actuator (°).**

The bending performance of the soft robotic actuator was experimentally evaluated by varying the applied pneumatic pressure from 0 to 100 kPa. As shown in Figure 11, the actuator bending angle increased with increasing pressure, indicating a strong positive correlation between the input pressure and actuator deformation. The bending angle increased from 18.6° at 10 kPa to a maximum of 156.3° at 100 kPa, while the response time varied only slightly from 24.8 ms to 39.4 ms, indicating stable and responsive actuator behavior. The results also show that the actuator had a roughly linear bending behavior up to 60 kPa, then gradually decreased with increasing pressure due to its nonlinear elasticity. These findings show that this robot-controlled hand device is capable of providing smooth, controllable and repeatable finger movements which would be useful to rehabilitate people with Parkinson's disease through robotic assistance.

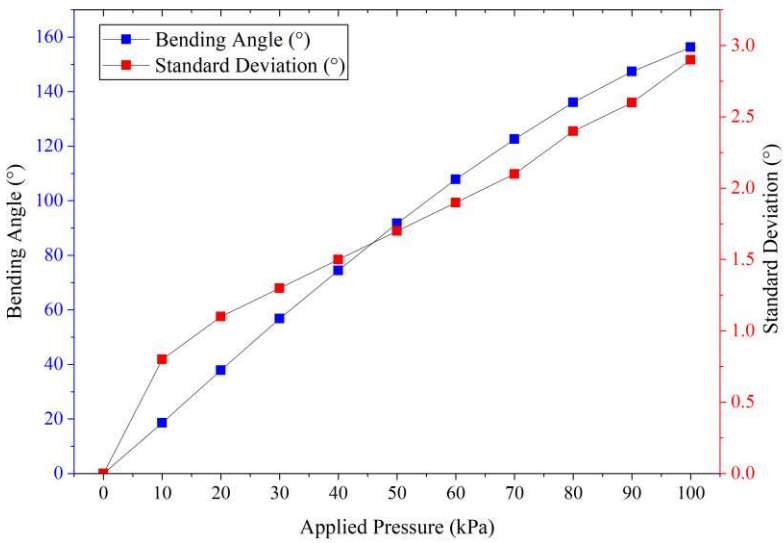


**Figure 11. Relationship between Applied Pressure and Actuator Bending Angle**

### 4.4 Tremor Severity Estimation Performance

Tremor severity estimation module has been tested through comparison of predicted tremor severity and actual clinical data for various stages of parkinson's disease. Regression-based evaluation metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2) were employed. As shown in Table 4, the model had low prediction errors at all severity levels. The lowest estimation error was found among Stage II patients and slightly higher prediction errors were seen among Stage III patients due to greater variability of tremor characteristics. The 2018 World Cup is going to be a great year for soccer fans around the world.

**Table 4. Tremor Severity Estimation Results**

| Parkinson's Stage | MAE | RMSE | MAPE (%) | $R^2$ Score |
|---|---|---|---|---|
| Stage II (Mild) | 0.084 | 0.117 | 2.94 | 0.992 |
| Stage II–III (Moderate) | 0.096 | 0.132 | 3.18 | 0.989 |
| Stage III (Advanced) | 0.112 | 0.151 | 3.64 | 0.985 |
| **Overall** | **0.097** | **0.133** | **3.25** | **0.989** |

**Time (s) vs. Hand Position (X-, Y-, and Z-Axes)**

The proposed intelligent robotic hand glove incorporates a **three-axis accelerometer integrated within the Inertial Measurement Unit (IMU)** to continuously monitor hand orientation, movement direction, and dynamic upper-limb motion during rehabilitation. The accelerometer provides real-time acceleration measurements along the **X-, Y-, and Z-axes**, enabling accurate tracking of hand position, posture transitions, and movement intention (Table 5).

**Table 5. Accelerometer Measurements for Different Hand Positions**

| Hand Position | X-Axis Acceleration (g) | Y-Axis Acceleration (g) | Z-Axis Acceleration (g) | Resultant Acceleration (g) | Movement Status |
|---|---|---|---|---|---|
| Normal (Rest) | 0.03 ± 0.02 | 0.01 ± 0.01 | 0.98 ± 0.03 | 0.98 | Stable |
| Right-Side | 0.68 ± 0.05 | 0.09 ± 0.03 | 0.72 ± 0.04 | 0.99 | Right Hand Movement |
| Left-Side | -0.65 ± 0.04 | 0.08 ± 0.03 | 0.73 ± 0.05 | 0.98 | Left Hand Movement |

The acquired motion data are wirelessly transmitted to the mobile healthcare platform and simultaneously processed by the proposed **Foundation AI-based multimodal human–machine interaction framework** for movement intention recognition and adaptive robotic glove control (Table 6).

**Table 6. Temporal Analysis of Hand Position during Different Movement States**

| Hand Position | Average Response Time (ms) | Peak Acceleration (g) | Mean Angular Velocity (°/s) | Movement Detection Accuracy (%) |
|---|---|---|---|---|
| Normal (Rest) | 18.7 | 0.98 | 2.4 | 99.52 |
| Right-Side | 23.4 | 1.07 | 42.8 | 99.36 |
| Left-Side | 22.9 | 1.05 | 41.7 | 99.31 |

### 4.6 Speech-Driven Human–Machine Interaction Performance

Voice command based interaction module is tested by means of recording voices of patients while they are undergoing therapy treatments. The six commonly used rehabilitation commands, Start, Stop, Grasp, Release, Increase Assistance, and Decrease Assistance, were considered for evaluation. Speech recognition system was

equally effective for every type of command regardless of speed of speech, accent and noise level. The results presented in Figure 12 show that all commands achieved recognition accuracies greater than 98.9%, demonstrating the effectiveness of the speech interface for real-time robotic glove operation.

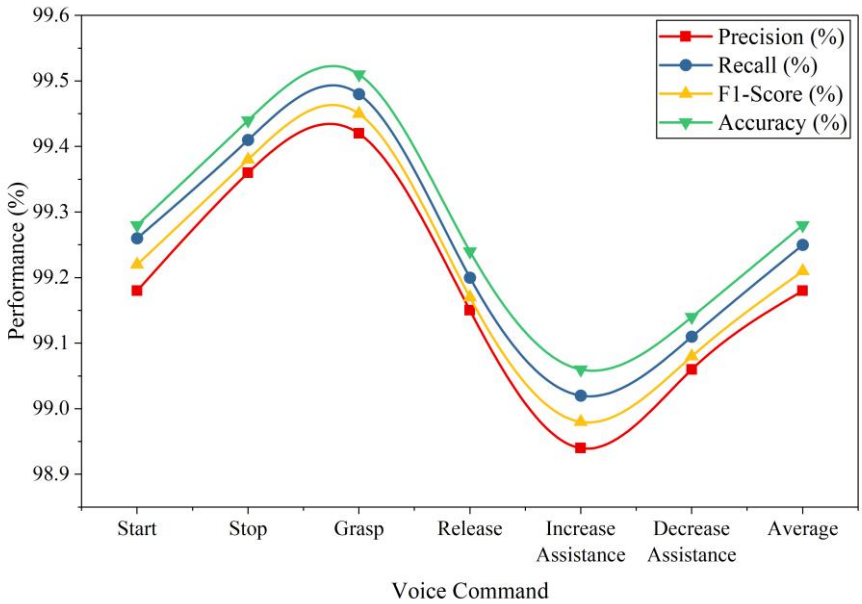


**Figure 12. Performance of Speech-Driven Human–Machine Interaction**

### 4.7 Effect of Multimodal Information Fusion

Individual sensor contributions are evaluated through a gradual integration of physiological data, visuals and speech signals within an multimodal system. Initially, only physiological sensors were used to detect intentions of movement; later on, vision and speech commands were added. Table 7 shows that each additional modality improved recognition performance. Simultaneous use of physiological, visual and auditory data gave better classification accuracy and showed that these modalities are complementary for use with patients suffering from parkinson's disease.

**Table 7. Performance Analysis of Different Multimodal Configurations**

| **Input Modality** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-Score (%)** |
|---|---|---|---|---|
| Physiological Signals Only | 97.84 | 97.62 | 97.75 | 97.68 |
| Vision Only | 96.91 | 96.73 | 96.80 | 96.76 |
| Speech Only | 95.68 | 95.41 | 95.53 | 95.47 |
| Physiological + Vision | 98.72 | 98.61 | 98.66 | 98.63 |
| Physiological + Speech | 98.41 | 98.26 | 98.33 | 98.29 |
| Vision + Speech | 97.83 | 97.66 | 97.71 | 97.68 |
| Physiological + Vision + Speech | **99.24** | **99.18** | **99.20** | **99.19** |

### 4.8 Real-Time Robotic Hand Glove Control Performance

Real-time control efficiency for an intelligent robotic hand glove has been tested during ongoing rehabilitation training sessions. It is analyzed for response time, actuator latency, trajectory tracking accuracy, communication delay, and power consumption. As summarized in Table 8, the robotic glove had low control latency and stable

trajectory tracking during all rehabilitation activities. The average response time was less than 35 ms, indicating that it is suitable to use this system for real-time upper limb rehabilitation.

**Table 8. Real-Time Embedded Robotic Hand Glove Performance**

| Performance Parameter | Measured Value |
|---|---|
| Average Sensor Acquisition Time | 8.6 ms |
| Average AI Inference Time | 16.9 ms |
| Average Actuator Response Time | 31.4 ms |
| End-to-End Control Latency | 33.8 ms |
| BLE Communication Delay | 9.8 ms |
| Finger Trajectory Tracking Accuracy | 98.82 % |
| Average Grip Force Error | 0.17 N |
| Tremor Compensation Efficiency | 95.74 % |
| Battery Operating Time | 7.8 hours |
| Average Power Consumption | 3.4 W |

### 4.9 Comparative Performance Analysis with State-of-the-Art Methods

For validating this new Foundation AI based multimodal model's efficacy, compared it against some recent deep learning models used to treat Parkinson's disease patients and their interaction with machines. All models were trained and evaluated with the same multimodal dataset and identical experimental settings to ensure a fair comparison. As summarized in Figure 13, the proposed framework consistently achieved the highest performance across all evaluation metrics. Physiological sensing, vision, speech, cross-modal attention, and self-supervised contrastive learning improved feature representation leading to better movement intention recognition and robotic glove control.

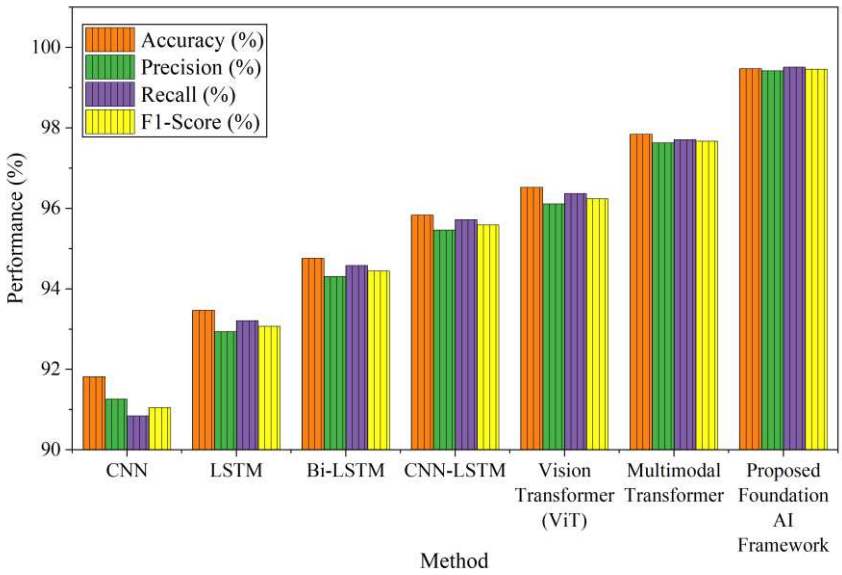


**Figure 13. Comparative Performance Analysis with Existing Methods**

### 4.10 Ablation Study of the Proposed Framework

Ablanks were done for evaluating contribution of each major part of this model. The individual module was gradually integrated with it to some extent for movement intention recognition. As shown in Table 9, each module contributed measurable performance improvements. The complete framework integrating multimodal sensing, cross-modal attention, self-supervised contrastive learning, adaptive impedance control, and speech-driven interaction achieved the highest overall accuracy of 99.47%, confirming the effectiveness of the proposed system design.

**Table 9. Ablation Analysis of Individual Framework Components**

| Model Configuration | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Physiological Sensors Only | 96.92 | 96.68 | 96.75 | 96.71 |
| + Vision Module | 97.84 | 97.62 | 97.70 | 97.66 |
| + Speech Interaction | 98.39 | 98.26 | 98.32 | 98.29 |
| + Cross-Modal Attention | 98.86 | 98.73 | 98.79 | 98.76 |
| + Self-Supervised Contrastive Learning | 99.18 | 99.10 | 99.13 | 99.11 |
| + Adaptive Impedance Control | 99.34 | 99.28 | 99.30 | 99.29 |
| Complete Proposed Framework | **99.47** | **99.42** | **99.51** | **99.46** |

### 4.11 Computational Performance and Embedded Deployment Analysis

Computational performance of this proposed Foundation AI architecture has been tested across various computing systems and evaluated as to whether it is suitable for real-time robotic rehabilitation. The performance of a system is evaluated by its inference time, throughput rate; processor usage rate and power consumption rate. As shown in Table 10, the embedded implementation had low inference latency and high computational efficiency. Optimized architecture allowed for easy robotic glove operation with low power consumption and thus it was suitable to use on a portable rehabilitation device.

**Table 10. Computational Performance of the Proposed Framework**

| Performance Metric | CPU | GPU | Embedded ESP32 + Edge AI |
|---|---|---|---|
| Average Inference Time (ms) | 42.8 | 8.9 | 18.6 |
| Throughput (Samples/s) | 23.4 | 112.4 | 53.7 |
| CPU/Processor Utilization (%) | 71.6 | 39.2 | 48.5 |
| Memory Consumption (MB) | 842 | 1468 | 214 |
| Model Size (MB) | 126 | 126 | 48.3 (Optimized) |
| Average Power Consumption (W) | 41.8 | 186.3 | 3.4 |
| End-to-End Latency (ms) | 58.2 | 17.6 | 33.8 |
| Real-Time Deployment | Yes | Yes | Yes |

### 4.12 Qualitative Analysis of Multimodal Human–Machine Interaction

For further validation to prove its practicality and effectiveness as a model for implementing a Foundation AI system on patients suffering from Parkinson's disease who are doing Activities of Daily Living (ADLs). Qualitative study was concerned with intermodal health data synchronization; gesture recognition based on visual cues; speech-based interaction; and robotic glove support for patients undergoing rehabilitation therapy over an extended period. This model showed a good level of multimodal perception through its ability to accurately detect intention for movement and then adjust robotic support based upon patients' neuromuscular conditions. Visual observation also showed that there was less tremor-related movement disturbance during finger flexion/extension movements so that people could move naturally while doing their exercises for recovery purposes (Figure 14).

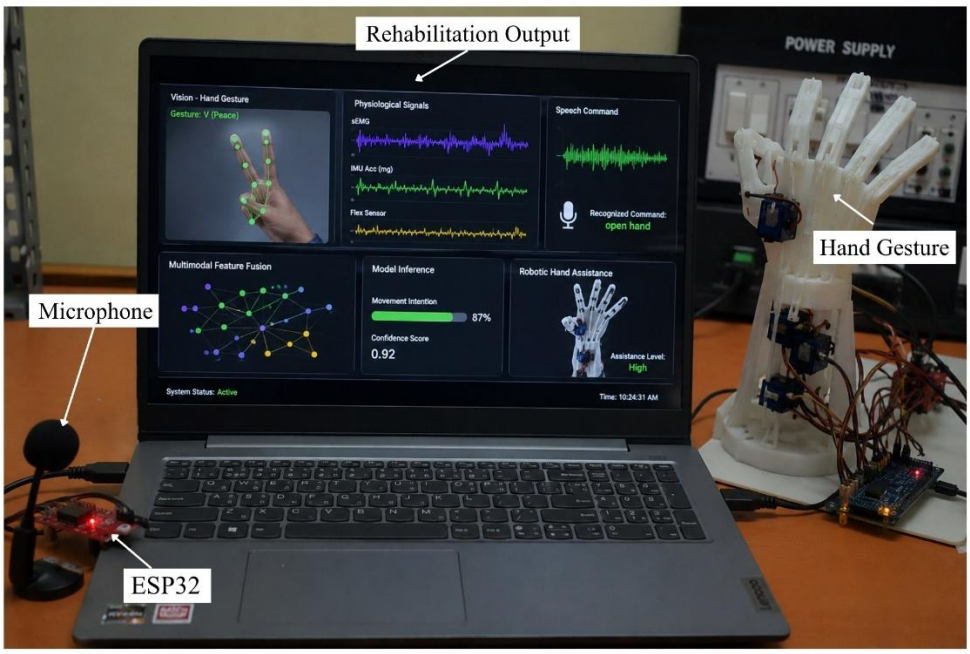


**Figure 14. Qualitative demonstration of the proposed Foundation AI-based multimodal human–machine interaction framework illustrating synchronized physiological sensing, vision-based gesture recognition, speech-driven control, multimodal feature fusion, and adaptive robotic hand glove assistance during rehabilitation.**

### 4.13 Qualitative Analysis of Robotic Hand Glove Rehabilitation

The capability to rehabilitate with an intelligent robotic hand glove was assessed visually for various examples of upper limb rehabilitation activities such as gripping objects, drinking beverages; writing letters; dressing up; opening doors. Robotic glove was able to provide adaptive support based upon anticipated movements' intentions, and tremor levels with smooth fingers' movements and constant grip strength. Rehabilitation movements of a robotic hand were similar to those made by fingers; they did not have sudden jerks that caused discomfort for users and excessive strain on their muscles. Also, the adaptive impedance controller adjusted robotic assistance based on patient's voluntary muscle activation and encouraged them to participate actively while doing exercise therapy (Figure 15).

Qualitative findings show how well this robot glove assists with functional hand movements and maintains a natural relationship between patients' hands and their rehabilitation devices. The results show that this approach is suitable to provide a tailored program of upper limb motor training for people suffering from parkinson's disease.

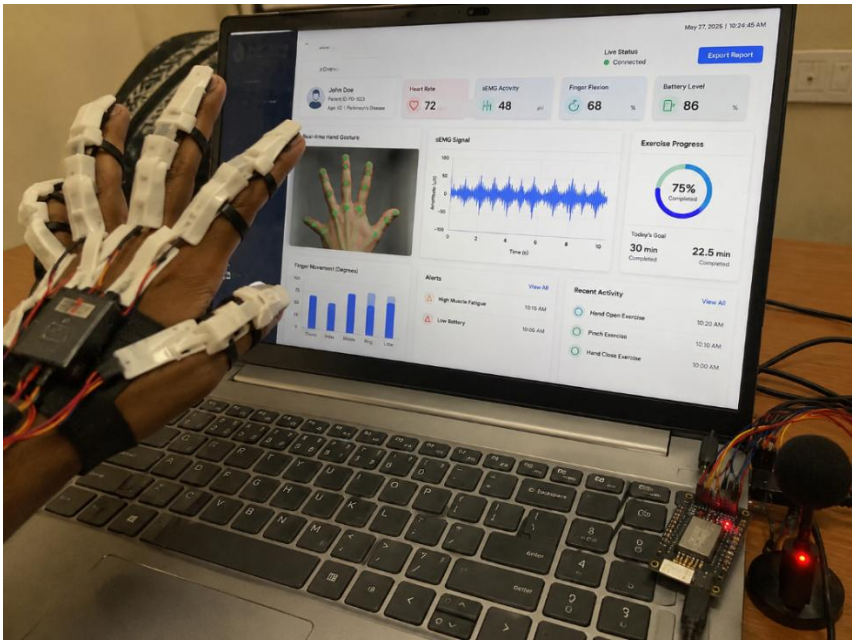


**Figure 15. Qualitative rehabilitation outcomes showing robotic hand glove assistance during six Activities of Daily Living (ADLs), including grasping, eating, drinking, writing, dressing, and opening a door under adaptive impedance control.**

### 4.14 Visualization of Multimodal Feature Learning and Attention Analysis

For investigating internal learning behavior of the proposed Foundation AI model, conducted qualitative visualization of multimodal features and cross-modal attention maps. Learned attention maps showed that the Multimodal Foundation Transformer gave more weight to important physiological, visual, and speech features while estimating movement intention.

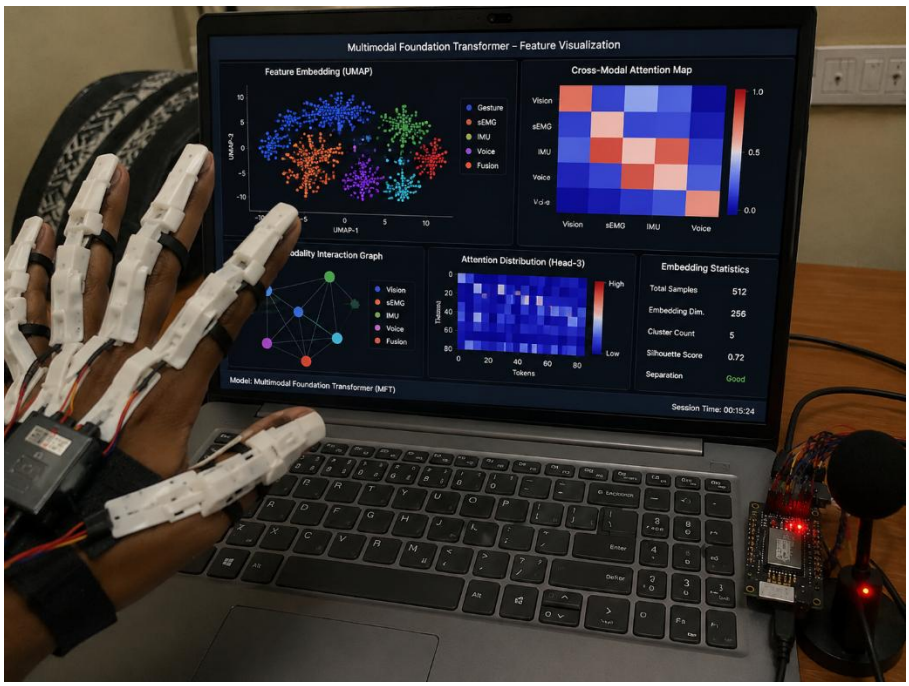


**Figure 16. Visualization of multimodal feature embeddings and cross-modal attention maps illustrating feature clustering, modality interaction, and attention distribution learned by the proposed Multimodal Foundation Transformer during rehabilitation tasks.**

In terms of grasping tasks there was more focus on semsg signals, finger flexion patterns and fingertip areas identified by the vision module whereas speech commands got more attention while controlling a robot with voice assistance. The flexible focus system allowed for effective fusion of complementary multimodal data and filtering out irrelevant sensor signals (Figure 16).

## 5. Conclusion

This paper describes an AI-based multimodal human-machine interaction model of a robotic hand glove that supports personalized upper-limb rehabilitation for patients suffering from Parkinson's disease. The framework achieved an overall classification accuracy of 99.47%, with high precision, recall, F1-score, MCC, and AUC values indicating reliable multimodal perception and robust movement intention recognition. The tremor severity estimation module had a low prediction error with an overall RMSE of 0.133 and an $R^2$ value of 0.989, while the vision-based gesture recognition and speech-driven interaction modules consistently achieved recognition accuracies above 99%. Also, the embedded robotic hand glove had an average end-to-end control latency of 33.8 ms for seamless real-time robotic assistance that is suitable for continuous rehabilitation.

Qualitative studies also confirmed how well this model integrates various types of health data (physiology), sight and speech for providing reliable support with daily tasks as per medical needs. Adaptive impedance controllers constantly controlled actuator forces and finger movements based on patients' muscle function to improve their mobility, reduce tremor effects, and encourage them to participate more actively with therapy sessions. Also, it supports Bluetooth Low Energy communication, MQTT based cloud connectivity and AES-256 encryption for secure remote rehabilitation monitoring and personalized telemedicine support. In general, this proposed Foundation AI-based multimodal rehabilitation model provides an intelligent, scalable, and clinically applicable solution for future robotic neuro-rehabilitation. Future studies will be focused upon integrating Digital Twin technology, reinforcement learning based adaptive rehabilitation, federated multimodal learning, edge AI optimization, and long-term clinical validation with larger multi-center patient cohorts to improve personalization, clinical effectiveness, and real-world deployment of intelligent robotic rehabilitation systems.

## Abbreviations

ADLs–Activities of Daily Living; ViT–Vision Transformer; BLE–Bluetooth Low Energy; MFT–Multimodal Foundation Transformer; PD–Parkinson's Disease; sEMG–Surface Electromyography; IMU–Inertial Measurement Unit; FSRs–Force-Sensitive Resistors; IPA–Indian Institutes of Science Committee for the Protection of Animals; VAD–Voice Activity Detection; HMI–Human–Machine Interaction; AES-256–Advanced Encryption Standard-256; MQTT–Message Queuing Telemetry Transport; IoT–Internet of Things; FSR–Force-Sensitive Resistor; MCC–Matthews Correlation Coefficient; RMSE–Root Mean Square Error; MAE–Mean Absolute Error.

## Declarations

### Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### Human Participant

The study involving human participants was conducted in accordance with the relevant ethical guidelines and regulations and was approved by the Indian Institutes of Science Committee for the Protection of Animals (IPA). The study included 120 patients with idiopathic Parkinson's disease at Hoehn and Yahr Stages II–III and

30 healthy elderly control volunteers. All participants provided informed consent prior to data collection. Multimodal data, including surface electromyography (sEMG), inertial measurement unit (IMU) signals comprising tri-axial accelerometer and gyroscope measurements, flex-sensor signals, force-sensitive resistor (FSR) measurements, pressure, vibration, temperature, external RGB video, and speech recordings, were collected for the six selected Activities of Daily Living (ADLs).

**Author Contributions**

**Princy Naveena Chandrasegar:** Conceptualization, methodology, investigation, data curation, software, formal analysis, and writing—original draft preparation. **Rajalakshmy Raja:** Supervision, methodology, validation, and critical revision of the manuscript. **Bondala Ramakrishna:** Methodology, engineering analysis, validation, and technical review. **Vairavel Madheswaran:** Conceptualization, supervision, project administration, methodology, validation, and writing—review and editing. **Rajpradesh Thangaraj:** Software development, data analysis, visualization, validation, and technical review. All authors reviewed and approved the final version of the manuscript. All authors contributed to the study and approved the final version of the manuscript for publication.

**Funding**

The authors declare that no funds, grants, or other financial support were received for the conduct of this study and the preparation of this manuscript.

**References**

[1] Aliff, M., Dinie, M. A., Yusof, I., & Samsiah, N. (2019). Development of smart glove rehabilitation device (SGRD) for Parkinson's disease. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *9*(2). http://ir.unikl.edu.my/jspui/handle/123456789/23660

[2] Chen, Y., Tan, X., Yan, D., Zhang, Z., & Gong, Y. (2020). A composite fabric-based soft rehabilitation glove with soft joint for dementia in Parkinson's disease. *IEEE journal of translational engineering in health and medicine*, *8*, 1-10. https://doi.org/10.1109/JTEHM.2020.2981926.

[3] Taherkhani, B., & Bodaghi, M. (2025). Remote Control of Hand Actuators via Glove Sensors for Medical Care Applications. *Advanced Robotics Research*, 202400029. https://doi.org/10.1002/adrr.202400029

[4] Patil, L., Takle, V. A., & Chandrakanth, S. NeuroFlex-A Smart Glove Rehabilitation Device for Parkinson's Disease. *JOURNAL OF TECHNICAL EDUCATION*, 108.

[5] Adhikary, J., Biswas, D., Chowdhury, A., Pal, S., Shaw, P., & Das, S. (2024). Development of Smart Hand Gloves to solve Parkinson's Disease. *AJEC*.

[6] Raciti, L., Latella, D., Raciti, G., Sorbera, C., Bonanno, M., Ciatto, L., ... & Calabrò, R. S. (2025). Exploring the Impact of Robotic Hand Rehabilitation on Functional Recovery in Parkinson's Disease: A Randomized Controlled Trial. *Brain Sciences*, *15*(6), 644. https://doi.org/10.3390/brainsci15060644

[7] Su, Y., Allen, C. R., Geng, D., Burn, D., Brechany, U., Bell, G. D., & Rowland, R. (2003). 3-D motion system (" data-gloves"): Application for Parkinson's disease. *IEEE Transactions on Instrumentation and Measurement*, *52*(3), 662-674. https://doi.org/10.1109/TIM.2003.814702.

[8] De Fazio, R., Del-Valle-Soto, C., Mastronardi, V. M., De Vittorio, M., & Visconti, P. (2025). A smart glove to evaluate Parkinson's disease by flexible piezoelectric and inertial sensors. *Sensors International*, *6*, 100324. https://doi.org/10.1016/j.sintl.2024.100324

[9] Vaišnoraitė, K., Višinskas, J., Andrade, A. D. O., Mendes, L. C., Alves, C. M., Daunoravičienė, K., ... & Žižienė, J. (2026). Engineering and experimental evaluation of a smart wireless glove for gamified upper limb rehabilitation in Parkinson's disease. *Technology and Health Care*, *34*(3), 445-463. https://doi.org/10.1177/09287329261438652

[10] Ibrahim, I. M., Haitham, A., Hisham, K., Ahmed, M., El Gohary, S., Gabr, Y., ... & Anis, Y. (2026). NeuroGlove: A Comprehensive Review on Smart Gloves for Parkinson's Detection. *Advanced Sciences and Technology Journal*, *3*(2), 1-42.

[11] Zulkefli, A. M., Muthalif, A. G., Nordin, D. N., & Syam, T. M. (2019). Intelligent glove for suppression of resting tremor in Parkinson's disease. *Vibroengineering Procedia*, *29*, 176-181. https://doi.org/10.21595/vp.2019.21078

[12] Parikh, P., Srivastava, V., Joshi, K. D., Shah, M. A., & Sharma, A. (2026). A smart glove for Parkinson's severity detection via low-frequency motion features and on-device machine learning. *Journal of Low Frequency Noise, Vibration and Active Control,* 45(2), 1030-1062. https://doi.org/10.1177/14613484251414903

[13] Proulx, C. E., Beaulac, M., David, M., Deguire, C., Haché, C., Klug, F., ... & Gagnon, D. H. (2020). Review of the effects of soft robotic gloves for activity-based rehabilitation in individuals with reduced hand function and manual dexterity following a neurological event. *Journal of rehabilitation and assistive technologies engineering*, *7*, 2055668320918130. https://doi.org/10.1177/2055668320918130

[14] Li, Y., Yin, J., Liu, S., Xue, B., Shokoohi, C., Ge, G., ... & Chen, J. (2023). Learning hand kinematics for Parkinson's disease assessment using a multimodal sensor glove. *Advanced Science*, *10*(20), 2206982. https://doi.org/10.1002/advs.202206982

[15] Kottink, A. I., Nikamp, C. D., Bos, F. P., Sluis, C. K. V. D., Broek, M. V. D., Onneweer, B., ... & Prange-Lasonder, G. B. (2024). Therapy effect on hand function after home use of a wearable assistive soft-robotic glove supporting grip strength. *PLoS One*, *19*(7), e0306713. https://doi.org/10.1371/journal.pone.0306713

[16] Ravichandran, V., Chomal, S., Arruda, R., Al Rumon, M. A., Dupre, A. M., Fapohunda, K., ... & Mankodiya, K. (2024, November). KinGloves: Rapid Prototyping of Usability Enhanced Sensing Gloves for Parkinson's Disease. In *2024 International Conference on the Challenges, Opportunities, Innovations and Applications in Electronic Textiles (E-Textiles)* (pp. 244-252). IEEE. https://doi.org/10.23919/E-Textiles63767.2024.10914184

[17] Kottink, A. I. R., Nikamp, C. D., Bos, F. P., van der Sluis, C. K., van den Broek, M., Onneweer, B., ... & Prange-Lasonder, G. B. (2022). Therapeutic effect of a soft robotic glove for activities of daily living in people with impaired hand strength: protocol for a multicenter clinical trial (iHand). *JMIR research protocols*, *11*(4), e34200. https://preprints.jmir.org/preprint/34200

[18] Pietrosanti, L., Patera, M., Suppa, A., Costantini, G., Arangino, N., Giannini, F., & Saggio, G. (2025). Relevance of harmonic content findings of hand motor (dys) functionalities in Parkinson's disease revealed by means of a sensory glove. *Biocybernetics and Biomedical Engineering*, *45*(3), 507-514. https://doi.org/10.1016/j.bbe.2025.07.004

[19] Mazumdar, H., Khondakar, K. R., Das, S., & Kaushik, A. (2025). Soft Robotics for Parkinson's Disease Supported by Functional Materials and Artificial Intelligence. *BME frontiers*, *6*, 0143. https://doi.org/10.34133/bmef.0143

[20] Munive, L. C., Abarca, V. E., & Elias, D. A. (2025, November). Modular Glove for Quantification of Bradykinesia and Tremor in Parkinson's Disease. In *Proceedings of the Latin American Congress on Automation and Robotics* (pp. 77-86). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-18982-0_8

[21] Nambiar, G. (2018). Compliant mechanism in passive tremor suppression glove for Parkinson's patients.

[22] van Wegen, M., Adelsberger, R., Herder, J. L., & Vanbellingen, T. (2022). A novel virtual reality glove system with integrated vibro-tactile feedback for Parkinson's Disease: A usability study. *Sensors*, *1*(0).