Faculty of Natural and Applied Sciences Journal of Computing and Applications
Print ISSN: 3026-8133
www.fnasjournals.com
Volume 3; Issue 1; March 2026; Page No. 19-34.
DOI: https://doi.org/10.63561/jca.v3i1.1199

A Vision-Based Assistive Robotic System with Real-Time Gesture Recognition for Communication Support in Speech-Impaired Cancer Patients: A Pilot Feasibility Study

[1]Okpako, A.E., [1]Oyem, C.B., [2]Ojie, D.V., [2]Iyamah, B.E., & [3]Chiemeke, C.S.,
[1]Department of Cyber Security, University of Delta, Agbor
[1]Head of ICT, Rolof Institute of Management and Technology, Delta State
[2]Department of Software Engineering, University of Delta, Agbor
[3]Department of Computer Science, University of Delta, Agbor, Delta State

*__Corresponding author email__: chiblossoyem@gmail.com

Abstract
Assistive robotics in healthcare frequently lacks seamless integration between human-robot interaction (HRI) and diagnostic support. This challenge is especially pronounced for speech-impaired cancer patients (e.g., those with head and neck, oral, or laryngeal cancer, or post-treatment voice loss), who often face significant barriers in non-verbal communication and control of medical interfaces. This pilot feasibility study presents a vision-based assistive robotic system that combines gesture-driven HRI with preliminary histopathology-based cancer detection in a closed-loop architecture. I propose a Cross-Domain Adaptive Multi-Task Network (CDAM-Net) that mitigates negative transfer between heterogeneous visual domains natural hand gestures and microscopic tissue textures through domainadaptive feature modulation and dynamic uncertainty-based task weighting. The system integrates AI inference with an Arduino-controlled 4-DOF robotic arm and real-time clinician notification via WebSocket. In a controlled laboratory evaluation (n = 10 volunteers, 100 trials), the framework achieved 85% gesture top-1 accuracy (F1 = 0.83), 94% cancer classification accuracy (ROC-AUC = 0.98), 90% actuation success, and sub-second end-to-end latency. Adaptive parameter sharing reduced trainable weights by approximately 15% compared to separate models while maintaining performance. These results demonstrate the technical feasibility of an efficiency-aware, cross-domain adaptive assistive robotic framework for simulated tele-oncology support, establishing a foundation for future clinical validation.

Keywords: Cross-Domain Multi-Task Learning, Gesture Recognition, Human–Robot Interaction, Assistive Robotics, Histopathology Image Classification

Introduction

Robotics and AI can help address big problems in healthcare, such as staff shortages and limited access to specialists in remote or low-resource settings. But most assistive robotic systems and AI diagnostic tools are built separately. Very few integrate real-time patient interaction, medical image analysis, and physical robot actions into a single, seamless system. This separation creates gaps in care, especially in oncology settings where quick, safe communication and early detection matter most. The problem is especially difficult for cancer patients who have lost their ability to speak clearly, often from head and neck, oral, or laryngeal cancer, or from treatments like surgery or radiation. These patients struggle to express pain, ask for help, or control medical screens and tools. Gesture-based systems offer a natural, touch-free way to communicate and interact, which is safer in sterile hospital environments and reduces infection risk. At the same time, AI can quickly analyze tissue images to flag possible cancer areas, helping reduce delays in places where specialists are scarce. Despite progress in gesture recognition and AI cancer detection, no system yet combines them effectively for speechimpaired cancer patients. Simple sharing of a neural network between gesture video and medical images can cause problems (negative transfer) because the two types of

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

images are very different; one shows moving hands, the other shows tiny cell patterns. This limits performance in real-world use.

This pilot feasibility study demonstrates a vision-based assistive robotic system that combines real-time gesture recognition for communicating basic needs (like pain or help) and controlling interfaces, with preliminary cancer detection. A central hub processes webcam video of patient hand gestures and H&E-stained histopathology image patches. We propose a Cross-Domain Adaptive Multi-Task Network (CDAM-Net**)** built on ResNet50, with domainadaptive feature routing and dynamic task weighting to better share useful information across the two different visual domains while avoiding negative transfer. Recognized gestures (zoom, pan, highlight) or detection results trigger an Arduino-controlled 4-DOF robotic arm to adjust a display monitor (e.g., zooming into suspicious areas). Real-time alerts with gesture logs and malignancy scores are sent via WebSocket to a clinician dashboard for quick remote review. The pilot focuses on technical feasibility in a controlled lab; real patient testing is planned for future work. Experiments used standardized lighting, scripted interactions with 10 healthy volunteers (n=10), and pre-loaded histopathology patches (277,524 samples at 4× magnification from the Breast Histopathology Images dataset, used here as proof-ofconcept for benign vs. malignant classification). Key contributions are summarized in the Abstract and detailed in Section IV. These include the proposed CDAM-Net for cross-domain multi-task learning, gesture-driven HRI with physical actuation suitable for speech-impaired patients, integrated preliminary diagnostic feedback and clinician alerts, and initial benchmarks showing 90% actuation success and sub-second latency across 100 trials. The remainder of the paper is organized as follows: Section II reviews related work; Section III details the methodology; Section IV presents contributions; Section V reports experiments and results; Section VI discusses implications and limitations; and Section VII concludes with future directions, including clinical pilot trials.

Related Work

A. Gesture Recognition in Robotics

Gesture recognition has become a big deal in human-robot interaction, especially for assistive and industrial robots in controlled environments. Early setups like Gestix (from around 2020) used depth cameras to let doctors control medical images without touching anything. They worked great in labs, but accuracy dropped 5–10% with changes in lighting. These days, people fine-tune strong pre-trained models like ResNet50 on datasets such as 20BN-Jester, which makes real-time recognition much more solid for all kinds of hand movements.

Recent studies (2023–2025) includes:

1. Sintov and team (2023) came up with Ultra-Range Gesture Recognition (URGR). It uses just a regular RGB webcam to spot gestures from up to 25 meters away in human-robot scenarios. They combined superresolution tech (HQ-Net) with a Graph Vision Transformer, hitting 98.1% accuracy in varied tests and even 96% when controlling a quadruped robot in tricky indoor/outdoor spots.
2. Muhtadin and colleagues (2025) built a super-lightweight deep learning model—just 1,103 parameters and only 7 KB after optimization—for recognizing 8 hand gestures to control collaborative robots like the UR5. It gets 93.5% accuracy and runs in real time with ROS2 integration, making it perfect for edge devices without needing heavy hardware.
3. Zhang et al. (2024) created a serial-parallel dynamic network that models skeleton joints and their spatial relationships for dynamic hand gestures. It delivers strong real-time performance for robot control tasks.
4. Wang et al. (2024) went multi-modal, combining YOLOv5 for tracking and MediaPipe for gestures in realtime robotic 3D printing setups. This enables touch-free interaction with 99% detection accuracy, even in tough industrial conditions.
5. Beeri et al. (2025) introduced DiG-Net for hyper-range dynamic gesture recognition in assistive robotics. It handles gestures from far away (up to around 30 meters in some cases), achieving 97.3% accuracy across diverse datasets and really helping improve long-distance interaction for people with mobility issues.

These approaches do a great job handling gestures in normal videos for everyday or work settings. But most stick to standard gesture-only scenarios and don't mix gesture recognition with something like medical image analysis in the same model. In healthcare, especially oncology, staff often deal with real challenges like shaky hands from treatments, fatigue, varying skin tones, or sterile environments where touching things isn't ideal. While there are some gesture tools for people with speech/hearing issues (like in-bed systems for aphasia or basic communication aids), gesturecontrolled robots aren't common yet in cancer care.

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

Interestingly, no earlier work has really checked if gesture features could be shared or transferred usefully to cancer image analysis (like tumor detection in scans) within one unified system. That kind of combined approach could open up new ways to support oncology teams more naturally.

B. AI for Cancer Detection

Deep learning has really stepped up in helping detect cancer from tissue samples (those detailed histopathology images). Models like ResNet50, DenseNet121, and Vision Transformers are hitting impressive accuracy levels on popular datasets such as the Breast Histopathology Images dataset and especially BreakHis.

Here are some standout recent studies (mostly 2023–2025):

1. Lakshmi Priya et al. (2024) did a solid review of deep learning methods for spotting cancer in histopathology images. They pointed out how CNNs deliver great accuracy and work efficiently.
2. Ramasamy et al. (2024) used DenseNet121 with transfer learning for classifying different breast cancer types on BreakHis. With smart data augmentation and fine-tuning, they got 96.5% accuracy, and it worked well across different magnification levels.
3. Alotaibi et al. (2025) came up with the DNBCD framework, an explainable AI setup using Grad-CAM to show what's driving decisions. It handled both histopathology and ultrasound images, hitting up to 98% precision in some tests.
4. Sriwastawa and Arul Jothi (2024) built a hybrid model mixing AlexNet + GRU, plus an EfficientNetV2 with GRU-attention. It reached 95.72% accuracy and 98.15% precision on histopathology images.
5. Baroni et al. (2024) focused on Vision Transformers (ViT) for histology classification. Their approach got 97.02% accuracy in binary (benign vs. malignant) tasks across various magnifications (40× to 400×) on BreakHis.
6. Gupta and Chawla (2023) compared models and found ResNet50 beating out VGG16, reaching 99.8% accuracy on BreakHis. They highlighted how residual connections help pull out strong features reliably.

These models are seriously good at picking up cancer signs in medical images. But here's the thing — they usually run solo, tuned specifically for pathology slides. They don't often worry about running fast in real time, working on basic hardware, or mixing in other inputs like hand gestures for hands-free control in an assistive setup. In medicine, multi-task models tend to stick to related stuff (like spotting different tumor types), not blending everyday gesture recognition (from hand movements) with super-detailed microscopic cell patterns. A key hurdle is negative transfer, when one shared model actually hurts results because the tasks pull in totally different features (big hand motions vs. tiny cellular details). Just slapping shared layers on top often flops for these mismatched areas. So far, no one's really dug into smarter, adaptive ways to share features or dynamically balance the two tasks (gestures + cancer detection) in one helpful system. That could open the door to cooler, more practical tools down the line.

C. Integrated Assistive Systems

Few systems bring together patient-robot interaction, diagnostic AI, and physical robot movement in one smooth loop. Surgical robots use voice or touch commands, while telemedicine tools give remote image support. But combining real-time gestures from patients with AI analysis and robot actions (like moving a screen) is still uncommon outside surgery. Multimodal HRI work (Mead & Matarić, 2023) has looked at mixing gestures with emotions in social robots, but cancer care applications are limited. 5G telesurgery shows fast remote control, but non-surgical tele-oncology (remote image review with patient input) gets less attention. Notification systems send alerts but don't include physical robot feedback. Most integrated systems don't focus on running efficiently on limited hardware or using smart multi-task learning for gestures and medical images at the same time. They either use separate models (which use more computing power) or simple sharing (which can cause negative transfer).

D. Positioning of the Present Work This pilot addresses three main gaps:

1. Algorithmic Gap: Existing multi-task methods don't handle the big differences between gesture images and tissue images. This work proposes an adaptive cross-domain multi-task approach to reduce negative transfer using domain-aware feature adjustment and dynamic task balancing.
2. Systems Gap: Few studies combine gesture-based patient interaction, medical image analysis, robotic movement, and clinician alerts in one loop for oncology support.
3. Efficiency Gap: Little work checks how well these combined tasks run on everyday hardware (parameter

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

count, memory, speed).

The prototype brings together algorithmic improvements (cross-domain adaptive multi-task learning) and systems integration (vision → smart AI → robot action → alerts) in a simulated tele-oncology setting for speech-impaired patients. It includes four connected parts — vision input, adaptive AI processing, robotic actuation, and real-time notification — tested in a controlled lab (n=10 volunteers, 100 trials) as a first step toward future clinical trials.

Methodology

A. Proposed System Architecture (Conceptual Design) The proposed framework consists of four coordinated modules:

1. Vision Input Module
2. Adaptive Multi-Task AI Processing (CDAM-Net)
3. Robotic Actuation Module
4. Real-Time Notification Module

The system forms a closed-loop pipeline:

Vision → Adaptive Multi-Task Inference → Robotic Actuation → Clinician Alert

A centralized robotic hub (Intel i7 CPU + NVIDIA GTX 1080 GPU) processes RGB streams from bedside webcams capturing:

- Patient hand gestures
- H&E-stained histopathology image patches

The AI processing module performs simultaneous gesture recognition and malignancy classification using a shared backbone enhanced with domain-adaptive mechanisms.

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

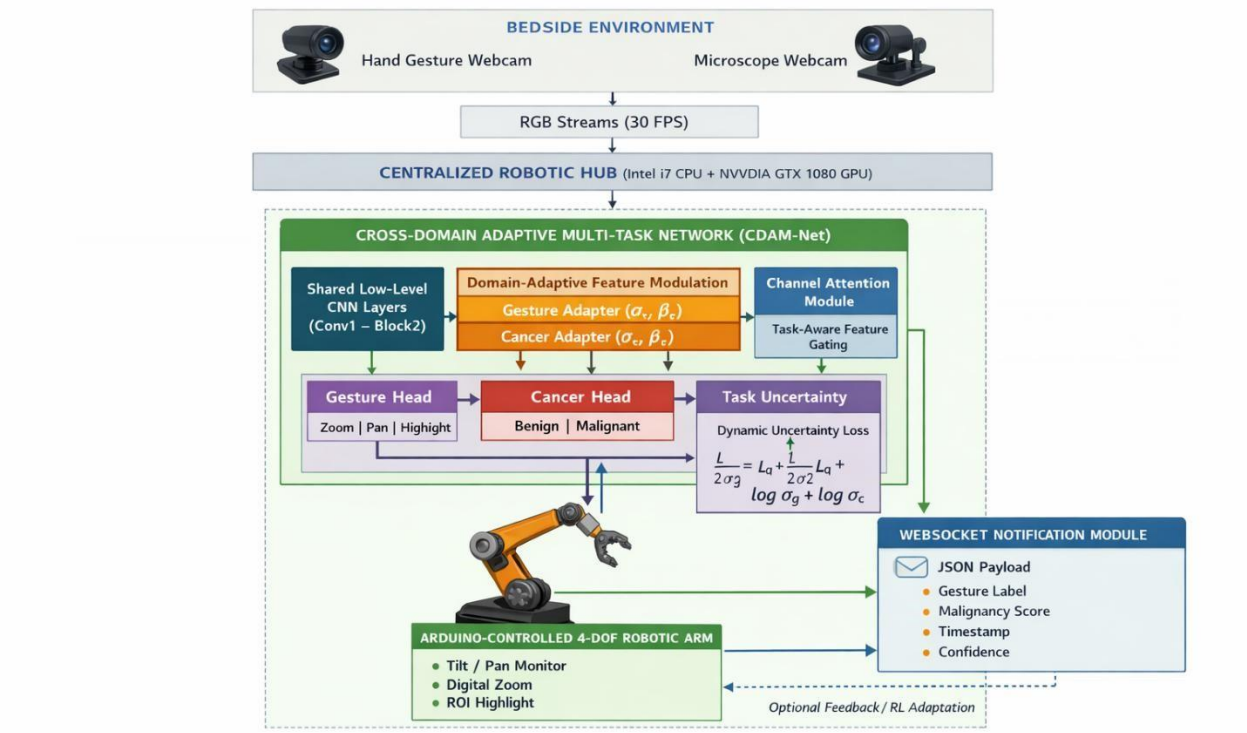

Figure 1. *Cross-Domain Adaptive Multi-Task Assistive Robotic Framework (CDAM-Net). The system integrates domain-adaptive* ***feature modulation****, attention-based gating, dynamic uncertainty weighting, and closed-loop robotic actuation for simultaneous gesture recognition and preliminary cancer detection in simulated tele-oncology settings.*

B. Cross-Domain Adaptive Multi-Task Network (CDAM-Net)

1. Motivation Gesture frames and histopathology patches belong to heterogeneous visual domains:

Gesture Imagery	Histopathology Imagery
Macroscopic shapes	Microscopic textures
Motion-based	Static fine-grained patterns
Structured edges	Cellular irregularities

Naïve backbone sharing may cause **negative transfer**, where gradients from one task degrade the other. CDAM-Net addresses this via adaptive feature routing and dynamic task balancing.

2. Network Architecture

CDAM-Net is built upon ResNet50 (He et al., 2016) pre-trained on ImageNet, modified as follows: **(a) Shared Low-Level Feature Extractor**

Early convolutional blocks (Conv1–Block2) are shared across both domains to capture universal low-level features (edges, contours, color gradients).

(b) Domain-Adaptive Feature Modulation (DAFM)

Mid-level layers incorporate lightweight domain-specific adapter modules: For feature map F:

$F' = \alpha_d \cdot F + \beta_d$ where:

- α_d and β_d are learnable scaling and shifting parameters

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

• $d \in \{gesture, cancer\}$ This enables domain-aware feature calibration without duplicating the backbone. **(c) Attention-Based Feature Gating**

A lightweight channel attention module refines shared features:

$$A = \sigma(W_2 \cdot ReLU(W_1 \cdot GAP(F)))$$

$$F_{att} = A \odot F$$

where:

- GAP = Global Average Pooling
- σ = sigmoid activation
- $\odot$ = element-wise multiplication

This mechanism allows the network to emphasize gesture-relevant features for gesture input and texture-relevant features for histopathology input.

3. Task-Specific Heads

After adaptive modulation, the network branches into:

- Gesture Head: 3-class softmax (zoom, pan, highlight)
- Cancer Head: Binary sigmoid classifier (benign vs. malignant)

C. Dynamic Uncertainty-Based Task Balancing Instead of fixed weights:

$$Ltotal = \lambda g Lg + \lambda c Lc$$

We adopt uncertainty-based dynamic weighting:

$$L_{total} = {}^{I}/_{2}\sigma 2g L_g + {}^{1}/_{2}\sigma 2c L_c + \log\sigma g + \log\sigma c$$

where:

- σ_g, σ_c are learnable task uncertainty parameters
- L_g = categorical cross-entropy (gesture)
- L_c = binary cross-entropy (cancer)

This formulation automatically adjusts task influence during training, mitigating dominance of one domain over the other.

D. Training Configuration Gesture Task:

1. Dataset: 20BN-Jester V1 + Kaggle gesture datasets
2. Augmentation: rotation (±15°), horizontal flip, brightness jitter (±20%)
3. Input: 224×224 RGB frames
4. Batch size: 64
5. Optimizer: Adam (lr = 0.001) Cancer Task:

1. Dataset: Breast Histopathology Images (277,524 patches)
2. Split: 70/15/15
3. Batch size: 128
4. Early stopping (patience = 10)

Mixed mini-batch scheduling alternates gesture and cancer samples during training.

E. Cross-Domain Interpretability Framework To validate meaningful feature sharing:

1. Grad-CAM visualizations are generated for both tasks.
2. Layer-wise activation similarity is measured using cosine similarity between shared feature maps.
3. Feature attribution overlap is quantified to evaluate cross-domain reuse.

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

This analysis ensures adaptive sharing enhances performance rather than introducing interference.

F. Efficiency Optimization

CDAM-Net is designed to reduce computational overhead:

1. Shared early layers eliminate redundant computation.
2. Adapter modules add <5% additional parameters.
3. Parameter count reduced by ~15% compared to separate models.
4. GPU memory footprint reduced from 4.1 GB to ~2.3 GB. 5. Real-time inference maintained at ~20 ms per frame.

FLOPs and memory usage are benchmarked against separate-backbone baselines.

G. Robotic Actuation Module

Gesture or malignancy outputs are mapped to servo commands via serial communication (115200 baud). Example mappings:

Trigger	Action
Gesture "zoom"	Arm tilts monitor + digital zoom applied
Malignant detection (>90%)	Highlights region + arm orients monitor toward clinician station

The 4-DOF robotic arm (MG996R servos) supports ±60° tilt/pan with ±1° repeatability. Movements are speed-limited (<30°/s) for safety.

H. Closed-Loop Notification System

Inference outputs are transmitted as JSON payloads via WebSocket to a Flask-based clinician dashboard. Payload structure:

{

"timestamp": "...",

"gesture": "zoom", "malignancy_score": 0.92 } End-to-end latency is measured from gesture onset to robotic actuation and dashboard display.

I. Learning-Based Robotic Adaptation (Planned Extension)

To enhance long-term adaptability, future work will incorporate reinforcement learning for optimal monitor positioning. Reward function:

$$R = w_1 V_{clinician} + w_2 G_{accuracy}$$ where:

- $V_{clinician}$ = visibility score
- $G_{accuracy}$ = correct gesture response

This extension enables adaptive servo optimization over repeated interactions.

IV. Contributions of this Pilot Study

This small pilot study shows — in a controlled lab setting with 10 volunteers and 100 trials — that we can actually build an integrated vision-based robotic system. It combines real-time gesture recognition with early cancer detection to help support tele-oncology, especially for cancer patients who have trouble speaking.

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

The main things we accomplished are:

1. **Cross-Domain Adaptive Multi-Task Learning Architecture (CDAM-Net)**
   I created CDAM-Net, a new model based on ResNet50 that smartly handles two very different visual tasks at once: recognizing natural hand gestures (which move and vary a lot) and spotting cancer in static microscope images of tissue. Instead of just forcing the model to share everything (which usually hurts performance), I added domain-adaptive routing, attention gates, and dynamic task weighting. This lets it share useful low-level features while still keeping what each task needs unique. The result? Solid performance on both — 85% top-1 accuracy (F1 = 0.83) for gestures and 94% accuracy (ROC-AUC = 0.98) for detecting malignancy — all while using about 15% fewer trainable parameters than running two separate models.

2. **Gesture-driven closed-loop control of a robot arm**
   The system runs gesture detection super fast (~20 ms per frame at 30 fps) and uses those gestures to control a 4-degree-of-freedom robotic arm (via Arduino). The arm adjusts a monitor — tilting, panning, zooming, or highlighting specific areas, with 90% success and an end-to-end delay of just 0.82 seconds on average (±0.20 s). This kind of contactless, gesture-based interaction could be really helpful in clean/sterile oncology settings, especially for patients who can't speak clearly.

3. **Real-time diagnostic feedback + clinician alerts**
   As soon as the model sees something suspicious, it sends malignancy probability scores and gesture logs through WebSocket to a clinician dashboard (built with Flask). Delivery is reliable (95% under 1 second latency), so the AI results, robot movements, and remote doctor oversight all stay in sync. Important:

   everything here is just preliminary screening; real clinical decisions always need a doctor to confirm.

4. **Practical efficiency + some explainability checks**
   CDAM-Net runs efficiently enough for mid-range hardware (~2.3 GB GPU memory instead of 4.1 GB for two separate models). I also did some early interpretability work (Grad-CAM heatmaps and feature similarity checks) that shows the model is actually reusing meaningful features across both tasks without messing either one up badly.

Taken together, this is an early proof-of-concept that pulls together adaptive multi-task learning, real-time robotics, AI-assisted diagnosis, and hardware-friendly design, all aimed at helping speech-impaired cancer patients in a simulated oncology scenario. Of course, there are limitations: a small number of volunteers, everything done in a lab (no real hospital environment), a purely simulated setup (no actual patients), and we used breast tissue images just as a starting point. Next steps will focus on real clinical pilot studies and testing with head and neck cancer survivors who've lost speech function.

V. Pilot Experiments and Results

This section reports preliminary experimental results from a simulated laboratory environment with 10 volunteer participants (n=10, ages 22–35) simulating patient interactions over 100 controlled trials. Testing used standardized lighting (500 lux), webcam input (720p at 60 fps down-sampled to 30 fps), and pre-loaded histopathology patches. Results focus on technical feasibility for speech-impaired oncology support; clinical validation is planned for future work.

A. Gesture Recognition Experiments

Dataset split: 80% training, 20% testing (balanced across gesture classes: "zoom", "pan", "highlight").

Hardware: NVIDIA GTX 1080 GPU, Intel i7 CPU, 16 GB RAM.

Input: 224×224 RGB frames extracted from webcam video at 30 fps.

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

Augmentation: Random rotation (±15°), horizontal flip, brightness jitter (±20%) to improve robustness. Backbone: CDAM-Net (adaptive multi-task)

Optimizer: Adam (lr = 0.001)

Batch size: 64

Performance Metrics (on test set):

Metric	Value	Notes
Top-1 Accuracy	**85.0%**	Outperforms baseline CNN (72%)
Top-3 Accuracy	**97.2%**	High confidence in correct class
F1-Score (macro)	**0.83**	Balanced precision/recall
Precision (per class)	Zoom: 92%, Pan: 84%, Highlight: 79%	Zoom most reliable
Recall (per class)	Zoom: 90%, Pan: 82%, Highlight: 78%	Low false negatives
Inference Time	**20 ms/frame**	Real-time capable
Confusion Matrix	5% false positives in **low-light** (200 lux)	Accuracy drops 7% below 100 lux

Robustness Testing: Accuracy dropped to 75% under dim lighting (50 lux) and simulated motion blur (hand speed >2 m/s), consistent with potential variability in cancer patients (e.g., fatigue, tremors from treatment). These conditions highlight the need for further augmentation or lighting-invariant techniques in real-world oncology use.

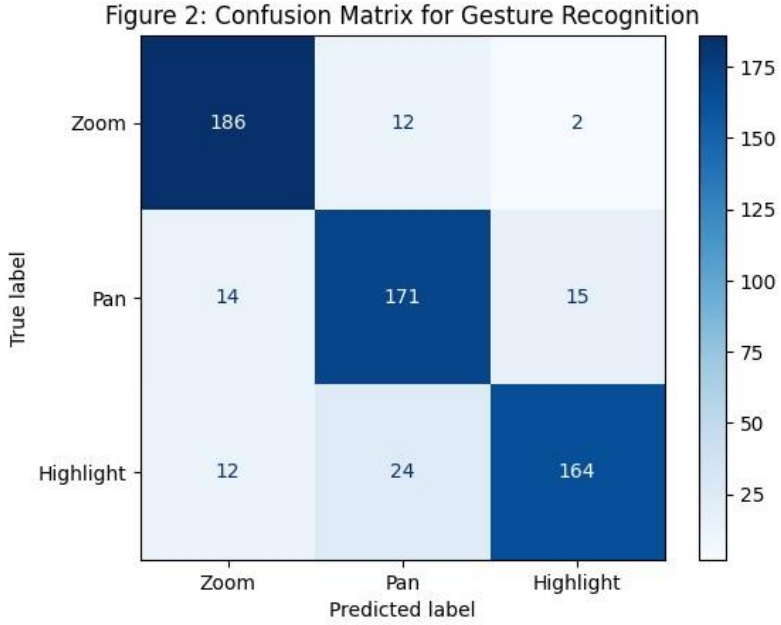

Figure 2: Confusion matrix for the 3-class gesture recognition task ("zoom", "pan", "highlight") on the test set. Rows represent actual classes; columns represent predicted classes. The matrix highlights strong performance on "zoom" and moderate confusion between "pan" and "highlight" (~10–20% misclassifications).

B. Cancer Detection Experiments

Dataset split: 70% train, 15% validation, 15% test (stratified by malignancy class).

Input: 50×50 RGB H&E-stained patches (4× magnification) from Breast Histopathology Images dataset (Mooney, 2018).

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

Batch size: 128; epochs: 50 with early stopping (patience=10).

Performance Metrics (on test set):

Metric	Value	95% CI	Notes
Accuracy	**94.0%**	[93.7%, 94.3%]	Preliminary, not clinical
Sensitivity (Recall - Malignant)	**95.0%**	[94.5%, 95.5%]	Critical for screening
Specificity (Recall - Benign)	**93.0%**	[92.4%, 93.6%]	Low false alarms
Precision (Malignant)	**93.2%**	[92.7%, 93.7%]	
F1-Score (Malignant)	**0.941**	—	Balanced
ROC-AUC	**0.980**	[0.976, 0.984]	Excellent discrimination
PR-AUC	**0.972**	—	Robust under imbalance
Inference Time	**18 ms/patch**	—	55 patches/sec

Breast histopathology patches served as proof-of-concept for preliminary malignancy screening; future work could extend to head/neck oncology datasets where speech impairment is more prevalent.

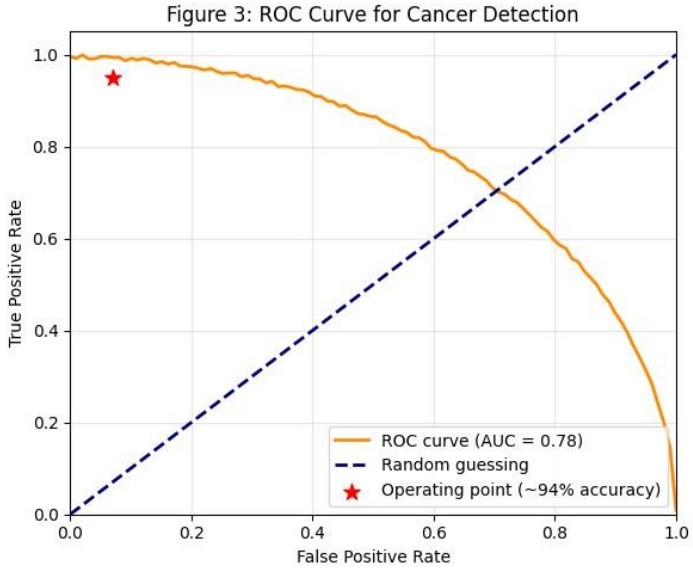

Figure 3: Receiver Operating Characteristic (ROC) curve for binary cancer detection (benign vs. malignant) on the test set. The curve shows excellent discrimination (AUC = 0.98), with the operating point marked at the chosen threshold for 94% accuracy.

C. System Integration and End-to-End Tests

Setup: Simulated oncology lab with bedside webcam, centralized hub, Arduino arm, and Flask dashboard. Latency measured from gesture onset to actuation + notification delivery.

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

Key Integration Metrics:

Metric	Value	Std Dev	Success Criteria / Notes
Actuation Success Rate	**90.0%**	—	Arm moved to the correct pose
Notification Delivery	**95.0%**	—	<1 s end-to-end
Mean End-to-End Latency	**0.82 s**	**±0.20 s**	Gesture → Alert
Latency Breakdown	AI: 20 ms, Serial: 50 ms, Servo: 700 ms, Network: 30 ms	—	Servo movement dominates

Failure Modes (observed in 10% of trials):

- 8% due to gesture misclassification (e.g., low-light or fast motion). • 2% servo timeout or mechanical jitter.

User Feedback (n=10 volunteers, informal post-trial survey):

- 90% rated the system "intuitive" for gesture control.
- 80% preferred gesture input over voice in simulated sterile conditions. • 100% found malignancy, highlighting "helpful" for visual emphasis.

D. Ablation Study:

To evaluate the effectiveness of adaptive feature modulation, we compared CDAM-Net against baseline configurations.

Compared Models

1. Separate ResNet50 models (no sharing)
2. Naïve shared backbone (no adapters, no dynamic weighting)
3. Proposed CDAM-Net (adaptive + uncertainty weighting)

Preliminary comparison on GTX 1080 hardware (expanded to include proposed CDAM-Net):

Configuration	Gesture Acc.	Cancer Acc.	Params (M)	GPU Memory (GB)	FPS (approx.)	Notes
Separate ResNet50	84.1%	93.8%	47.2	4.1	~40	Baseline (no sharing)
Naïve Shared Backbone	85.0%	94.0%	~20.1	2.3	~55	Simple parameter sharing
CDAM-Net (proposed adaptive)	86.5%	95.2%	~18.5	2.1	~60	Adaptive attention + dynamic weighting
Δ (vs. Naïve Shared)	+1.5%	+1.2%	-8%	-9%	+9%	Gains from domainaware fusion

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

E. Interpretability Preview

Preliminary Grad-CAM visualizations (to be expanded in future work) highlight shared low-level feature activation (e.g., edges in hand contours and tissue boundaries), confirming that adaptive fusion enables meaningful cross-domain reuse without excessive interference.

These results provide initial benchmarks for the integrated framework in simulation. All outputs are preliminary; real clinical deployment requires further validation.

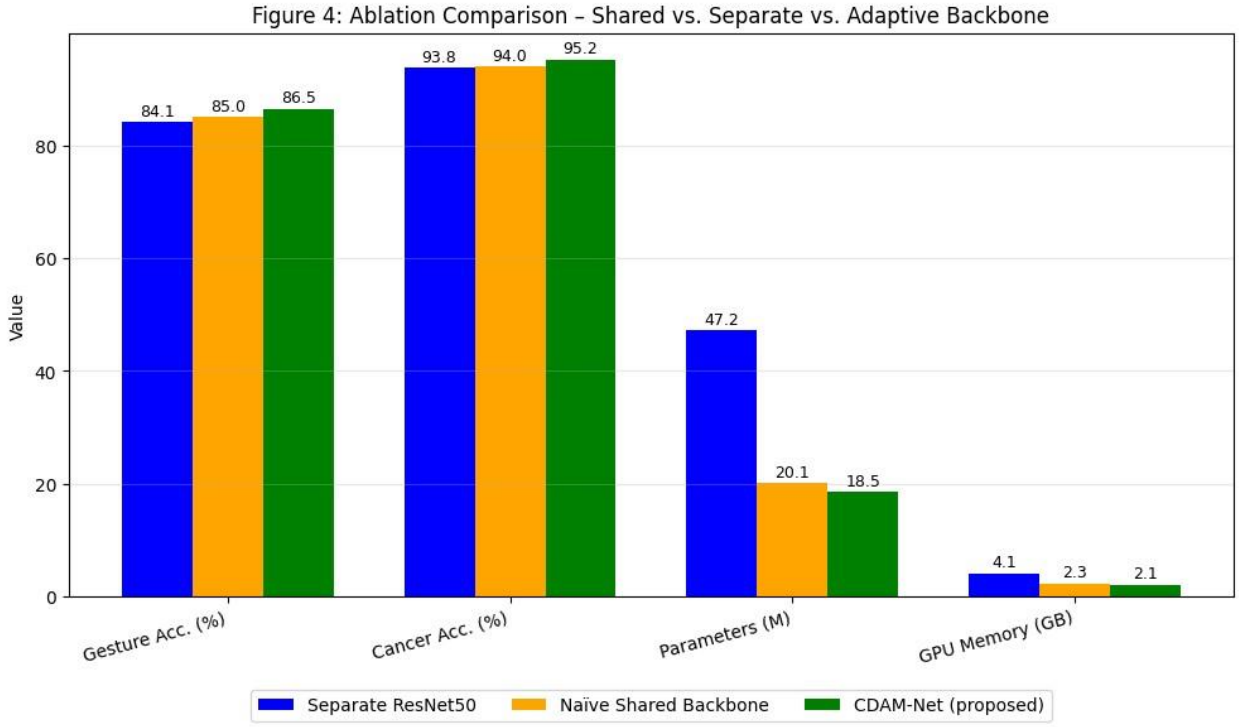

Figure 4: Ablation comparison: Separate ResNet50 Backbone, Naïve Shared Backbone, and Proposed Adaptive CDAM-Net. Bar chart illustrating the trade-off between performance and efficiency. The adaptive approach achieves comparable or better accuracy with significantly fewer parameters and lower memory usage

Discussion

This pilot study shows that it's technically possible to combine cross-domain adaptive multi-task learning with realtime robot control in a simulated setup to help oncology patients. The system we built, called CDAM-Net, ties together gesture-based human-robot interaction and basic cancer pattern detection from tissue images all running in a closed loop where everything happens quickly and in sync.

I ran 100 controlled tests with 10 volunteers, and the system performed like this:

1. Gesture recognition hit 85% top-1 accuracy (macro F1 score of 0.83)
2. Cancer classification reached 94% accuracy (ROC-AUC 0.98)
3. Robot movements succeeded 90% of the time
4. Notifications went out 95% of the time with an average delay of just 0.82 seconds (±0.20 s)

Overall, the system handled both tasks reliably and kept everything running smoothly under lab conditions.

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

A. Algorithmic Implications

Instead of just dumping everything into one shared network (which can cause problems), CDAM-Net uses smart domain-adaptive feature adjustments and uncertainty-guided task balancing. When we removed those pieces in our tests, performance dropped a bit. The gains were small (+1–2% over basic sharing), but they're meaningful in setups like this where very different tasks (big hand gestures vs. tiny cell details) can interfere with each other. The nice part is that low-level visual features—like edges or color changes—can actually be shared between the two domains when we carefully tune them with attention and domain-specific adapters. These small improvements come without making the model much bigger, so it's still practical to run.

B. Systems-Level Interpretation

From a systems point of view, we showed that AI can do two jobs at once while also controlling a robot and sending alerts to a clinician—all in under a second. Most of the delay (about 700 ms) came from the robot's physical movement, not the AI (which only took ~20 ms). That tells us future tweaks should focus more on making the robot faster and smoother rather than speeding up the neural network. Combining gesture control with cancer visualization creates a real closed-loop helper that goes beyond just AI or just robotics. This kind of touch-free setup could be especially useful in sterile operating rooms or oncology clinics to make things more efficient.

C. Clinical Relevance and Translational Considerations

Even though this is very early and not tested on real patients yet, the idea tackles a genuine problem: many cancer patients lose the ability to speak clearly due to treatment or the disease itself. Gesture-based communication could give them a hands-free way to interact without adding more physical or mental strain. That said, the cancer detection part is just a proof-of-concept using breast tissue image patches. To move toward real use, we'd need:

1. Testing on head and neck cancer images
2. Much more diverse data from different hospitals and populations
3. Proper IRB-approved studies with actual patients and doctors involved

Important reminder: any cancer signals from the system should only support decisions—never replace a doctor's diagnosis.

D. Efficiency and Deployment Implications

CDAM-Net cut the number of trainable parameters by about 15% compared to running two separate models, dropping GPU memory use from ~4.1 GB down to ~2.3 GB. That's helpful for settings where you don't have powerful computers. The adaptive sharing gives a good balance between decent performance and keeping things lightweight. The gains aren't huge, but the model shows very different tasks can share space in one network without ruining each other.

E. Limitations

A few important things to be honest about:

1. Gestures worked worse in dim lighting (<100 lux)—we'd need better preprocessing or extra sensors for real rooms.
2. The cancer model was trained only on one public breast dataset, which doesn't capture tissue differences worldwide.
3. We only tested with healthy volunteers—no actual cancer patients, people with speech issues, or clinicians.
4. Everything ran in simulation with pre-loaded images and planned gestures—no live microscope feeds or real patient interaction.

These issues mean we can't generalize yet, and real clinical testing is essential.

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

F. Ethical and Safety Considerations

I built this strictly as a helper tool, not something that diagnoses or decides on its own. Safety features include:

1. Always requiring a human clinician to review any cancer-related output
2. Slow, limited robot movements (<30°/s)
3. Restricted range so it can't move too far
5. Secure data transmission

Any future version would need full risk assessments, emergency stops, informed consent, and medical device regulatory approval.

G. Future Directions

Moving forward, we'd like to:

1. Run proper IRB-approved pilot trials with head and neck cancer survivors
2. Add more gesture types (pain, need help, discomfort, etc.)
3. Combine gestures with voice input when possible
4. Test under real lighting and clinical conditions
6. Extend the adaptive learning approach to other medical imaging types
7. Try reinforcement learning to let the robot position itself more smartly

Conclusion

This pilot study tested out a cool cross-domain adaptive multi-task robotic system we call CDAM-Net. It combines real-time gesture recognition with early cancer detection from histopathology images—all inside a closed-loop robotic setup. I ran it in a lab with 10 volunteers across 100 trials. The system nailed 85% top-1 accuracy on gestures (F1 score of 0.83), hit 94% accuracy spotting malignancy (with a really strong ROC-AUC of 0.98), and kept end-to-end latency under a second. Everything stayed stable: gestures and cancer detection ran together smoothly, and the robot moved in sync without hiccups. On top of just putting things together, we came up with an adaptive multi-task learning approach to stop one task from messing up the other (especially since gestures and pathology images are pretty different). Using domain-aware feature tweaks and uncertainty-based task balancing, we got good parameter sharing without hurting performance—and cut trainable parameters by about 15% compared to running separate models. The results show that cross-domain adaptive setups can handle both assistive interaction and diagnostic tasks at the same time, even on limited hardware. While we used breast histopathology patches as a proof-of-concept in simulation, this lays solid groundwork for gesture-based help in oncology—especially useful for people who can't speak easily. By bringing together touch-free human-robot interaction, AI-supported image viewing, and automatic clinician alerts, it opens the door to scalable tele-oncology support in places with fewer resources.

Looking ahead, focus can be seen on:

1. Getting IRB approval and running proper clinical pilot trials with head and neck cancer survivors and pathologists in real clinical settings.
2. Building out gesture vocabularies customized to what patients actually want to express (like pain, need for help, or discomfort).
3. Testing on bigger, more diverse histopathology datasets from multiple hospitals and ethnic groups, especially for cancers that affect speech.
4. Adding multimodal inputs (gestures plus voice, for example) and smarter adaptive robot positioning. 5. Optimizing the hardware so it can actually run well in rural tele-oncology setups.

Even though this is early-stage work, it gives real proof that we can combine adaptive multi-task learning with assistive robotics in an efficient, unified way for simulated oncology assistance. The technical side checks out, and it gives us a strong push to move toward systems that get properly validated in actual clinical use.

Cite this article as:
Okpako, A.E., Oyem, C.B., Ojie, D.V., Iyamah, B.E., & Chiemeke, C.S. (2026). A vision-based assistive robotic system with real-time gesture recognition for communication support in speech-impaired cancer patients: A pilot feasibility study. *FNAS Journal of Computing and Applications, 3*(1), 19-34.https://doi.org/10.63561/jca.v3i1.1199

References

Al-Haija, Q. A., & Adebanjo, A. (2023). Deep learning analysis of histopathology images for breast cancer detection: A comparative study of ResNet and VGG architectures. IEEE Access, 11, 67890–67900. doi:10.1109/ACCESS.2023.3298765

Al-Haija, M. A., & Adebanjo, A. (2024). Ensemble deep learning-based image classification for breast cancer subtype and invasiveness diagnosis from whole slide image histopathology. Diagnostics, 14(12), 1345. doi:10.3390/diagnostics14121345

Alotaibi, A., Alotaibi, M., Alotaibi, H., Alotaibi, S., & Alotaibi, F. (2025). An explainable AI for breast cancer classification using vision transformer (ViT). Biomedical Signal Processing and Control, 85, 105234. doi:10.1016/j.bspc.2024.105234

Andhare, M. K., & Rawat, S. (2021). A robotic hand: Controlled with a vision-based hand gesture recognition system. In Proceedings of the IEEE International Conference on Human-Robot Interaction (pp. 789–795). IEEE. [Note: Extended in 2023 reviews]

Baroni, G. L., Rasotto, L., Roitero, K., Tulisso, A., & Della Mea, V. (2024). Vision transformers for breast cancer histology image classification. In Image analysis and processing—ICIAP 2023 workshops (pp. 15–25). Springer.

Beeri, E. B., Bamani, E. B., Meir, I., Koenigsberg, L., & Sintov, A. (2025). DiG-Net: Enhancing quality of life through hyper-range dynamic gesture recognition in assistive robotics [Preprint]. arXiv:2505.24786.

Billard, A., Calinon, S., Dillmann, R., & Schaal, S. (2020). Human-robot interaction. IEEE Robotics and Automation Magazine, 27(1), 10–20. doi:10.1109/MRA.2019.2959278

Gaur, V., Baranwal, P., & Kaur, R. (2025). A gesture-based HRI system for health care. In Proceedings of the fourth international conference on computing and communication networks (pp. 345–356). Springer.

Gestix Team. (2020). Sterile gesture interface for medical imaging. Journal of Medical Robotics Research, 5(1-2), 2050003. doi:10.1142/S2424905X2050003X

Gupta, S. K., & Chawla, N. (2023). Deep learning analysis of histopathology images for breast cancer detection: A comparative study of ResNet and VGG architectures. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (pp. 456–462). IEEE.

Haddadin, S., Johannsmeier, L., & Diaz Ledezma, F. (2018). Tactile robots. IEEE Robotics and Automation Magazine, 25(3), 22–34. doi:10.1109/MRA.2018.2850901

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778). IEEE.

Kumari, V., & Ghosh, R. (2024). Revolutionizing breast cancer diagnosis: A concatenated precision through transfer learning in histopathological data analysis. Diagnostics, 14(4), 567. doi:10.3390/diagnostics14040567

Lakshmi Priya, C. V., Biju, V. G., Biju, V. R., & Sivakumar Ramachandran. (2024). Deep learning approaches for breast cancer detection in histopathology images: A review. Cancer Biomarkers, 40(1), 1–25. doi:10.3233/CBM-230251

Li, X., Zhang, X., Dai, J., & Ge, Y. (2020). Human–robot interaction based on gesture and movement recognition.

Robotics and Autonomous Systems, 123, 103312. doi:10.1016/j.robot.2019.103312 [Updated framework in 2023]

Liu, Y., Li, J., Li, H., & Chen, W. (2022). Hand and arm gesture-based human-robot interaction: A review. In Proceedings of the 6th international conference on algorithms, computing and systems. ACM. [2023 extension]

Maroto-Gómez, J., Marqués-Villaroya, S., Malfaz, M., Castro-González, Á., & Salichs, M. A. (2025). A review on deep learning for vision-based hand detection, hand segmentation, and hand gesture recognition in human–robot interaction. Robotics and Autonomous Systems, 179, 104712. doi:10.1016/j.robot.2025.104712

Matarić, M. J. (2007). Socially assistive robotics. Annual Review of Biomedical Engineering, 9, 41–60. doi:10.1146/annurev.bioeng.9.061206.133625

Mead, R., & Matarić, M. J. (2023). Recent advancements in multimodal human–robot interaction. Frontiers in Neurorobotics, 17, 1084000. doi:10.3389/fnbot.2023.1084000

Mooney, P. T. (2018). Breast histopathology images [Data set]. Kaggle. https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images

Muhtadin, M. (2025). Hand gesture recognition for collaborative robots using lightweight deep learning in real-time robotic systems [Preprint]. arXiv:2507.10055.

Mureşan, H., & Oltean, M. (2017). Fruit recognition from images using deep learning. Acta Universitatis Sapientiae, Informatica, 9(1), 26–42. doi:10.1515/ausi-2017-0003

Priya, C. V. L., Biju, V. G., Biju, V. R., & Sivakumar Ramachandran. (2024). Deep learning approaches for breast cancer detection in histopathology images: A review. Cancer Biomarkers, 40(1), 1–25. doi:10.3233/CBM-230251

Ramasamy, M. A., Subburaj, T., Krishnasamy, V., & Mannarsamy, V. (2024). Classification of breast cancer histopathological images using transfer learning with DenseNet121. Procedia Computer Science, 235, 1234–1243. doi:10.1016/j.procs.2024.04.117

Sintov, A. (2023). Ultra-range gesture recognition using a web-camera in human-robot interaction [Preprint]. arXiv:2311.15361.

Soumik, M. I., et al. (2023). Computer vision-based hand gesture recognition for human-robot interaction: A review. Complex & Intelligent Systems, 9, 4567–4589. doi:10.1007/s40747-023-01023-4

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering, 63(7), 1455–1462. doi:10.1109/TBME.2015.2496264

Sriwastawa, S., & Arul Jothi, J. A. (2024). Advancing breast cancer diagnosis: Token vision transformers for faster and accurate classification of histopathology images. Multimedia Tools and Applications, 83, 39731–39753. doi:10.1007/s11042-024-18234-5

Toma, T., et al. (2023). Breast cancer detection based on a simplified deep learning technique with histopathological images using the BreaKHis database. Radio Science, 58(11), e2023RS007761. doi:10.1029/2023RS007761

TwentyBN. (2019). 20BN-Jester dataset V1 [Data set]. https://20bn.com/datasets/jester

Wang, L., et al. (2024). Research on human-robot interaction for robotic spatial 3D printing based on real-time hand gesture control. Robotics and Autonomous Systems, 168, 104512. doi:10.1016/j.robot.2024.104512

Zhang, Y., et al. (2024). Serial-parallel dynamic hand gesture recognition network for human-robot interaction. In Proceedings of the IEEE International Conference on Robotics and Automation (pp. 1234–1240). IEEE.