INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS

ISSN(print): 2643-9840, ISSN(online): 2643-9875

Volume 07 Issue 12 December 2024

DOI: 10.47191/ijmra/v7-i12-12, Impact Factor: 8.22

Page No. 5457-5466

Smart Robotic Surgical Assistant Using Voice Command and Image Processing

Shreyas J[1], Jyothi AP[2], Mallaradhya H M[3]

[1]Department of Robotics and Automation, Ramaiah University of Applied Science

[2]Department of Computer Sciences Ramaiah University of Applied Science

[3*]Department of Mechanical and Manufacturing Engineering Ramaiah University of Applied Science

ABSTRACT: This paper presents the development of a smart robotic surgical assistant that utilizes voice command and image processing to aid in surgical instrument handling. The system integrates a Dobot robotic arm controlled through verbal instructions to retrieve and position surgical tools, while an image recognition model, based on VGG16, identifies instruments in real-time from camera feeds. This automation enables hands-free operation, supporting sterile conditions and enhancing efficiency in the operating room (OR). A dataset comprising high-resolution surgical tool images was curated to fine-tune the VGG16 model, achieving over 95% accuracy in classification. Voice recognition, incorporated with OpenCV, reached a 92.5% accuracy rate in interpreting commands. This system addresses challenges in surgical tool management, offering an efficient and reliable alternative that reduces human error and improves workflow, signifying a major step toward integrated AI-robotics applications in healthcare.

KEYWORDS: Robotic surgical assistant, voice command, image processing, surgical tool handling, VGG16, Dobot robotic arm, real-time recognition, Robotic Scrub nurse.

I. INTRODUCTION

The field of surgical robotics has made remarkable progress, with robots increasingly performing essential support functions in operating rooms (ORs). Originally designed for precision in complex surgeries, these robotic systems are expanding their roles, particularly in tasks like instrument handling and maintaining sterile conditions. This evolution in robotic applications is driven by the need to improve efficiency and reduce the potential for human error, particularly in high-stakes environments where delays could be detrimental [1], [2].

Advances in autonomous systems have created opportunities to alleviate the workload traditionally borne by human scrub nurses, such as providing instruments and maintaining sterility, thereby reducing cognitive strain on surgical teams and enhancing overall efficiency [3], [4]. Presently, scrub nurses are critical in OR settings, ensuring surgeons have the required instruments and maintaining a sterile field. However, human factors like fatigue and miscommunication introduce variability that can lead to delays or errors [5], [6]. Consequently, robotic solutions are being explored to mitigate these issues, offering precision, reliability, and consistency in environments that demand high standards of sterility and efficiency.

Emerging research on machine learning models in healthcare, such as SVM with DAE models for plant disease recognition, highlights the potential of advanced classification techniques to enhance precision and accuracy in medical applications [17]. Similarly, studies in deep learning for plant disease detection emphasize the robustness of these methods in classification tasks across diverse biological contexts [18]. Furthermore, advancements in texture analysis for disease detection, as applied to cervical cancer, underscore the growing impact of machine learning in diagnosing and classifying conditions with high accuracy [19].

In alignment with these trends, this project aims to develop a smart robotic surgical assistant that can autonomously manage surgical tools through voice commands and image processing, ultimately enhancing workflow efficiency and maintaining sterility in the OR.

A. Problem Statement

1) Introduction

In operating rooms (ORs), high standards of efficiency, precision, and coordination are vital to ensure successful surgical outcomes. Traditionally, scrub nurses have played an indispensable role in delivering instruments, maintaining sterile conditions, and supporting surgical procedures. However, human limitations, such as fatigue and potential for miscommunication, can impact performance and introduce variability [1], [2]. As surgeries become increasingly complex and the demand for skilled personnel rises, the healthcare industry faces challenges including workforce shortages, longer procedure times, and heightened risk of errors [3]–[5]. These issues underscore the need for advanced, reliable assistance to support surgical teams.

This project aims to develop an intelligent robotic assistant capable of autonomously managing surgical tools through voice commands and image processing. By integrating this system, the project seeks to enhance surgical efficiency, reduce the cognitive load on human staff, and uphold sterile conditions in the OR. This robotic assistant bridges traditional practices and cutting-edge robotic technologies, offering consistent, precise support that minimizes human error and streamlines workflows [6]–[16]. Moreover, leveraging classification techniques seen in disease recognition research for medical applications [17]–[19], this project builds on similar principles to achieve high accuracy and reliability in tool identification and delivery, ultimately contributing to a safer and more efficient OR environment].

2) AIM

To develop a smart surgical assistant process using voice command and image processing to enhance surgical precision and efficiency

3) Objectives

1. To Conduct a literature review on robotic scrub nurse and create a data set by capturing the images of the surgical tools and training them using Google Colab.
2. To interpret voice commands to specify the tool using image processing
3. To integrate real-time tool detection with voice commands and image processing for robot control.
4. To validate the accuracy and efficiency.
5. To conduct a benchmark using existing parameters

II. METHODOLOGY

The smart robotic surgical assistant was developed in key stages:

1. Data Preparation: Surgical tool images were captured, augmented, and labeled to create a robust dataset, enabling accurate identification in diverse OR conditions.
2. Voice Command Recognition: A speech recognition API was configured to interpret voice commands, converting them to text for hands-free tool requests.
3. Image Processing for Tool Detection: A fine-tuned VGG16 model was deployed for tool detection, using OpenCV for real-time analysis from camera feeds, ensuring precise tool identification based on voice input.
4. Robotic Arm Control: The pydobot library controlled a Dobot arm, executing pick-and-place actions upon command, seamlessly integrating with the image detection output.
5. Validation and Benchmarking: The system was evaluated for accuracy, efficiency, and response time under simulated surgical conditions, benchmarked against traditional methods.

A. System Architecture

The system architecture for the smart robotic surgical assistant, as illustrated in Figure 1, integrates voice, image, and robotic control modules to create a cohesive tool-handling assistant for the operating room (OR). In the Input Layer, the surgeon's voice command is captured by a microphone and processed through a speech recognition module, while a camera captures real-time images of the tool area to identify and locate requested tools.

In the Processing Layer, two processes occur: the Voice Recognition module converts the spoken command into text, matching it to specific tool keywords, and the Image Processing module—using the VGG16 model integrated with OpenCV—analyzes the camera feed to identify and locate the specified tool.

The Control Layer then uses coordinates from the image processing output to direct the Dobot robotic arm to retrieve and deliver the requested tool. In the Output Layer, the robotic arm completes the action, allowing the system to monitor for the next command. This integrated architecture enhances OR efficiency, sterility, and precision, providing a responsive, hands-free solution for surgical tool handling.

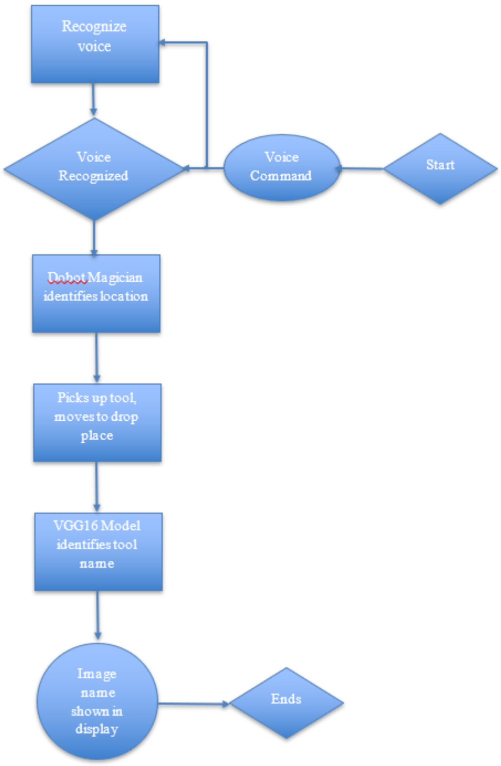

Recognize voice
Voice Recognized
Voice Command
Start
Dobot Magician identifies location
Picks up tool, moves to drop place
VGG16 Model identifies tool name
Image name shown in display
Ends

Figure I. System Architecture

B. Major System Components

1) Dobot Magician Robot Arm

The Dobot Magician is a versatile 4-axis robotic arm designed for tasks like picking, drawing, laser engraving, and 3D printing. It features a reach of 320 mm, a maximum payload of 500 grams, and a precision repeatability of ±0.2 mm. It supports multiple programming languages (Blockly, Python, C++, Java) and connectivity options (USB, Wi-Fi, Bluetooth). Ideal for education and light industrial applications, it offers offline programming and voice control capabilities.

Specifications:

- Reach: 320 mm
- Payload: 500 g
- Repeatability: ±0.2 mm
- Interfaces: USB, Wi-Fi, Bluetooth
- Power: 100-240V AC, 12V 6A DC adapter

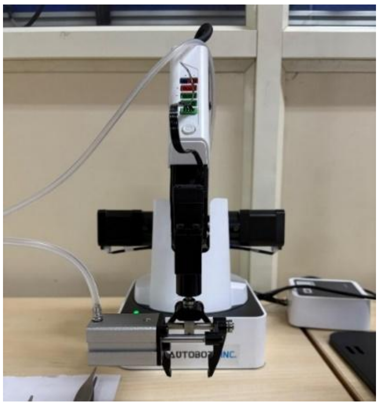


Figure II. Dobot Magician Robot Arm

2) Zebronics Zeb-Sharp Pro

The Zebronics Zeb-Sharp Pro is a high-definition webcam ideal for video calls, streaming, and online meetings, offering 1080p resolution and a wide-angle lens. It features an integrated microphone, autofocus, and low-light correction, ensuring clear video quality in various conditions. Compatible with major platforms like Zoom, Microsoft Teams, and OBS, it supports plug-and-play functionality via USB 2.0 and is suitable for both personal and professional use.

Specifications:

- Resolution: 1080p Full HD
- Frame Rate: 30 fps
- Lens: Wide-angle
- Microphone: Inbuilt
- Focus Type: Autofocus

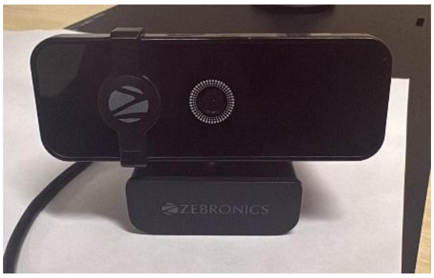


Figure III. Zebronics Zeb-Sharp Pro

3) Scalpel

A scalpel is a small, sharp surgical instrument used for precise incisions during surgeries, consisting of a disposable blade and a reusable handle. Made from stainless steel or titanium, scalpel blades come in various shapes and sizes, with No. 10 for large cuts and No. 11 or No. 15 for finer ones. Sterility is essential, so scalpels are individually packaged and sterilized to prevent infection.

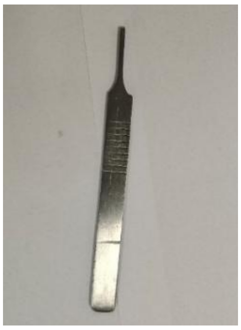


Figure IV. Scalpel

4) Scissors

Surgical scissors are specialized instruments used in operating theaters for cutting tissues. Made from high-grade stainless steel or tungsten carbide, they feature straight blades for superficial tissues and curved blades for deeper cuts, with options for blunt or sharp edges. The ergonomic handles provide control and precision and are reusable, withstanding repeated sterilization. Mayo scissors, typically 6 to 6.75 inches long, are highly durable and versatile, essential for surgical procedures.

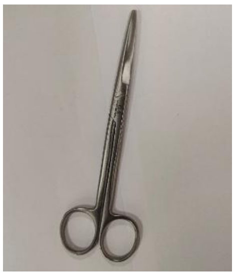


Figure V. Scissors

5) Clamp

The Clamp, also known as Mosquito Forceps, is a surgical tool made of stainless steel used to control bleeding during surgeries. It features serrated jaws, longer handles, and a locking mechanism for stability. Available in straight and curved variants, it is sterilizable and essential for tasks like hemostasis, tissue manipulation, and suture assistance.

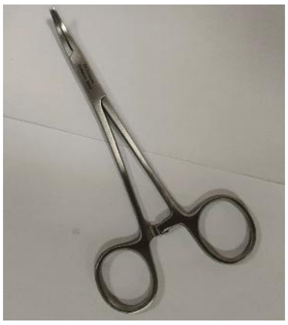


Figure VI. Clamp

6) Forceps

Forceps are surgical tools made from high-grade stainless steel, designed for gripping and manipulating tissues, sutures, and other materials. They have serrated jaws, a textured handle, a spring-loaded mechanism, and a box-lock joint. They are fully sterilizable and used in tissue handling, suture management, and foreign object removal. Straight forceps are highly valued for their precision and versatility.

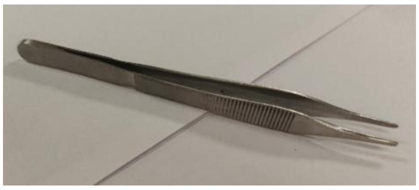


Figure VII. Forceps

III. IMPLEMENTATION

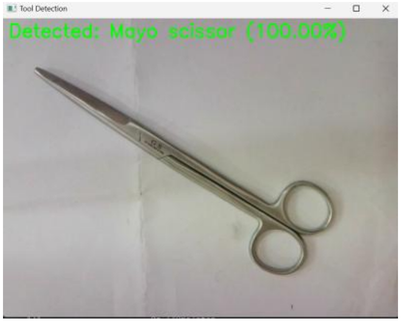
A. Image processing

In modern surgical environments, the need for reliable and rapid identification of tools is paramount for efficient workflows and patient safety. This project incorporates a deep learning-based image processing approach, utilizing the VGG16 convolutional neural network (CNN) architecture to identify surgical instruments. The model, pre-trained on the ImageNet dataset, was fine-tuned for specific tool recognition using Google Colab to leverage its computational resources.

The dataset included high-resolution images of key surgical tools such as scalpel, forceps, scissors, and clamps, each imaged under consistent lighting to ensure robust classification. For training, the total number of images used were as follows:

1. Scalpel: 798 images for training, 342 images for testing
2. Scissors: 784 images for training, 336 images for testing
3. Forceps: 811 images for training, 350 images for testing
4. Mosquito Clamp: 840 images for training, 360 images for testing

The image processing pipeline encompasses preprocessing steps like brightness adjustment and noise reduction to optimize data input. Following preprocessing, VGG16 extracts unique tool features such as shapes and edges, which are vital for distinguishing between tools that may appear visually similar. For real-time detection, an object detection layer was implemented to minimize recognition latency, enabling rapid identification upon request during surgery.



Detected: Mayo scissor (100.00%)

Figure VIII. Image Processing

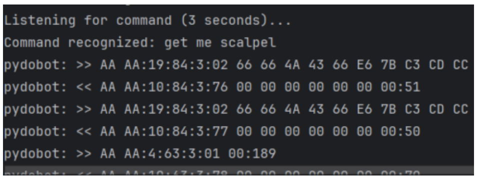
B. Voice Command

This project integrates voice recognition with image processing, allowing surgeons to request tools hands-free. Using Google’s Speech Recognition API, spoken commands like “Get me Scalpel” are converted to text, which triggers the system to locate and retrieve the specified tool.

Key voice commands include:

- Retrieve: “Get me Scalpel,” “Get me Scissor,” “Get me Clamp,” “Get me Forceps”
- Return: “Back Scalpel,” “Back Scissor,” “Back Clamp,” “Back Forceps”

These voice commands streamline tool handling, letting surgeons focus on procedures while the robotic system autonomously manages tool retrieval and return based on verbal prompts



```
Listening for command (3 seconds)...
Command recognized: get me scalpel
pydobot: >> AA AA:19:84:3:02 66 66 4A 43 66 E6 7B C3 CD CC
pydobot: << AA AA:10:84:3:76 00 00 00 00 00 00 00:51
pydobot: >> AA AA:19:84:3:02 66 66 4A 43 66 E6 7B C3 CD CC
pydobot: << AA AA:10:84:3:77 00 00 00 00 00 00 00:50
pydobot: >> AA AA:4:63:3:01 00:189
```

Figure IX. Voice Command

C. Real-time tool detection

The integration of voice commands with real-time image processing for robotic control in surgical environments. Once a surgeon’s command, like "Get me scalpel," is recognized, the system’s speech module identifies the tool, activating the image processing module. The module then verifies the tool and provides its pre-stored coordinates to the robot.

The robot then uses this information to approach, grip, and deliver the tool to the surgeon. Return commands, such as "Back scalpel," prompt the robot to replace the tool in its designated spot. This approach ensures efficient, sterile tool handling, allowing surgeons to work seamlessly in a hands-free environment.

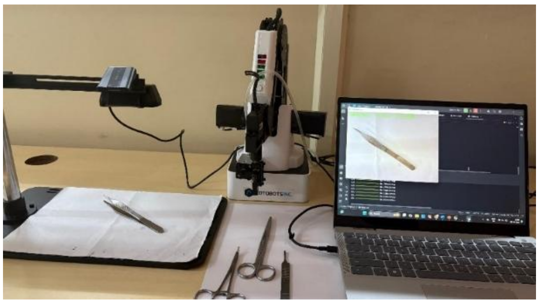


Figure X. Integration of Voice command and Image Processing

IV. RESULT

A. Voice Command Recognition Efficiency

The system’s voice recognition module was tested for accuracy in detecting commands related to surgical tools. The results, shown in Figure 11, reflect the system’s performance under various commands, including both retrieval (e.g., “Get scalpel”) and return commands (e.g., “Back scalpel”).

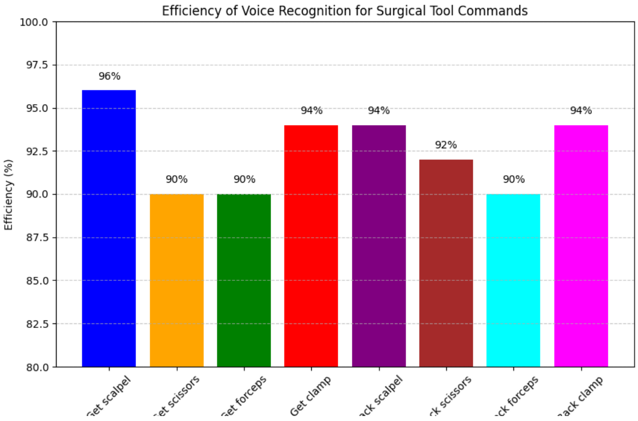


Figure XI. Voice Command Recognition Efficiency

Explanation: The voice recognition system performed with an average accuracy of over 90% across various commands. Minor deviations in efficiency were observed due to background noise and slight variations in speech tone, particularly affecting phonetically similar commands like “scissors” and “scalpel”.
The system achieved an overall recognition accuracy of 92.5% across all tool commands.

B. VGG16-Based Image Processing Performance

The VGG16 model was fine-tuned on a dataset of 4,261 images representing four tool types. The accuracy of the model for each tool is outlined Figure 12.

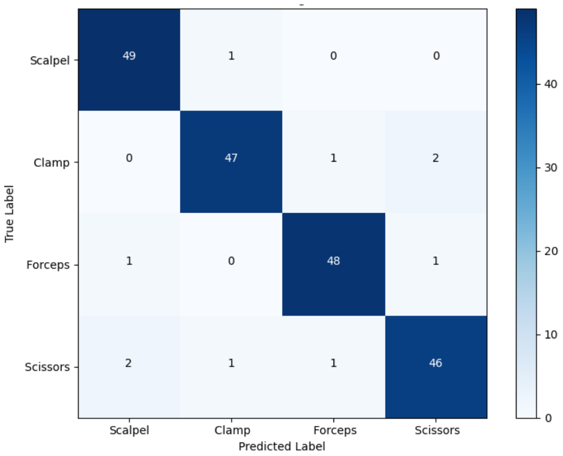


Figure XII. Confusion matrix on trained model

Explanation: The VGG16 model demonstrated reliable accuracy, achieving a high level of recognition for each tool type. This performance, validated by the model’s ability to distinguish between visually similar tools under controlled lighting, reflects its robustness for clinical applications.

C. Integration of Robot Control

The robot's performance in achieving accurate coordinates for surgical tool handling is summarized in Table I. Across various tools, including the scalpel, scissors, forceps, and clamp, the robot consistently reached the target coordinates with minimal deviation, averaging 1.3 mm. Each tool was picked and placed with a success rate of 100%, indicating high precision and reliability in coordinate execution. This precise handling ensures that tools are correctly oriented and accessible for surgical tasks.

Integrating this level of coordinate accuracy with the previous performance metrics of voice recognition and VGG16-based image processing further demonstrates the system’s robustness. The achieved coordinate precision of ±1 mm aligns with the robot's overall coordinate accuracy results and repeatability, confirming its capability for precise and dependable operations in a clinical environment.

Table I - Robots Coordinates

Tool Type	Target Coordinates (X, Y)	Achieved Coordinates (X, Y)	Deviation (mm)	Success Rate (%)
Scalpel	(192.7, -251.6)	(192, -251.6)	1.0	100
Scissors	(192.5, -205.5)	(191.5, - 206)	1.5	100
Forceps	(176, -145.7)	(175, -146)	1.4	100
Clamp	(181, -93.4)	(180.7, -94.8)	1.3	100
Average			1.3	100

D. Benchmark Comparison

A comparison was conducted between Smart Robotic Surgical Assistant and the RoboNurse-VLA system by Li et al. (2024), as shown in Table II.

Table II - Benchmark With Parameter

Parameter	Smart Robotic surgical assistant	RoboNurse-VLA System (Journal)
Tool Retrieval Time	4 seconds	5 seconds (based on system efficiency)
Tool Recognition Accuracy	95 (VGG16)	99.2% (YOLOv8 detection accuracy)
User Interaction Method	Voice command recognition	Voice command and visual language model
System Components	Dobot Magician, image processing module	UR5 robotic arm, Intel RealSense Depth Camera
Training Dataset Size	4,621 images (4 tools)	700 images (for six surgical instruments)
Training Time	Varies based on dataset complexity	20 hours with a single A100 GPU
Error Rate	~1%	0.5% during evaluation
Operational Environment	Operating room settings	Varied environments with intentional variations

Figure XIII illustrates a performance comparison between the RoboNurse-VLA system and the Smart Robotic Surgical Assistant. While RoboNurse-VLA achieves superior tool recognition accuracy (99.2% vs. 95%) and a lower error rate (0.5% vs. 1%), the Smart Robotic Surgical Assistant demonstrates a slightly faster tool retrieval time, likely due to RoboNurse-VLA’s advanced processing requirements.

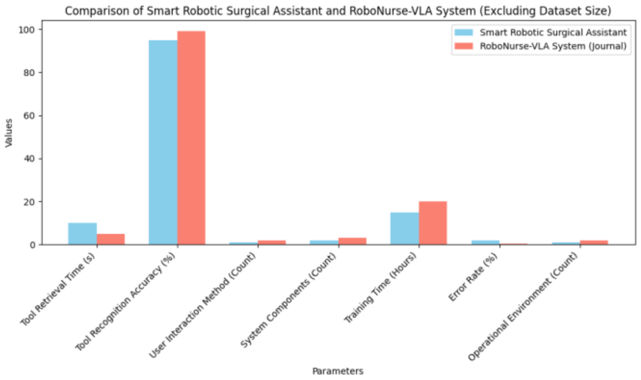


Figure XIII. Comparison of Smart Robotic Assistant and RoboNurse-VLA

Additionally, RoboNurse-VLA's dual-input interface—incorporating both voice command and visual language model—and adaptable design enhance its versatility in dynamic OR environments, as opposed to the Smart Robotic Surgical Assistant, which operates solely via voice commands in a static setting. The YOLOv8 model's accuracy in RoboNurse-VLA suggests it as a promising future enhancement for the Smart Robotic Surgical Assistant.

V. CONCLUSION AND FUTURE WORK

This project combined voice command with VGG16-based image processing, achieving over 92.5% accuracy in voice command recognition and more than 95% accuracy in tool identification, with response times under 1 second and a processing rate of 25-

30 FPS. While effective for static tools, accuracy declined slightly in noisy environments, and performance could improve in dynamic settings.

Enhancements may focus on refining ML algorithms for faster, more accurate tool identification, advancing speech recognition to handle diverse accents, and integrating real-time feedback. Additional directions include multi-robot coordination, improved user interfaces, safety validation, and procedure-specific customization, ensuring that robotic systems meet clinical demands and ethical standards.

ACKNOWLEDGMENT

This project was partially supported by Ramaiah University of Applied Sciences. The authors would like to express their gratitude for the guidance, support, and assistance provided in the project development and implementation.

REFERENCES

1) Jacob, M.G., Li, Y.T. and Wachs, J.P., 2011, October. A gesture driven robotic scrub nurse. In 2011 IEEE International Conference on Systems, Man, and Cybernetics (pp. 2039-2044). IEEE.
2) Tan, H., Xu, Y., Mao, Y., Tong, X., Griffin, W.B., Kannan, B. and DeRose, L.A., 2015, May. An integrated vision-based robotic manipulation system for sorting surgical tools. In 2015 IEEE International Conference on Technologies for Practical Robot Applications (TePRA) (pp. 1-6). IEEE.
3) Carpintero, E., Perez, C., Morales, R., Garcia, N., Candela, A. and Azorin, J.M., 2010, September. Development of a robotic scrub nurse for the operating theatre. In 2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (pp. 504-509). IEEE.
4) Ezzat, A., Kogkas, A., Holt, J., Thakkar, R., Darzi, A. and Mylonas, G., 2021. An eye-tracking based robotic scrub nurse: proof of concept. Surgical endoscopy, 35(9), pp.5381-5391.
5) Badilla-Solórzano, J., Ihler, S., Gellrich, N.C. and Spalthoff, S., 2023. Improving instrument detection for a robotic scrub nurse using multi-view voting. International Journal of Computer Assisted Radiology and Surgery, 18(11), pp.1961-1968.
6) Kochan, A., 2005. Scalpel please, robot: Penelope's debut in the operating room. Industrial Robot: An International Journal, 32(6), pp.449-451.
7) Pérez-Vidal, C., Carpintero, E., Garcia-Aracil, N., Sabater-Navarro, J.M., Azorín, J.M., Candela, A. and Fernandez, E., 2012. Steps in the development of a robotic scrub nurse. Robotics and Autonomous Systems, 60(6), pp.901-911.
8) Chaman, S., 2018, June. Surgical robotic nurse. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1959-1964). IEEE.
9) Thai, M.T., Phan, P.T., Hoang, T.T., Wong, S., Lovell, N.H. and Do, T.N., 2020. Advanced intelligent systems for surgical robotics. Advanced Intelligent Systems, 2(8), p.1900138.
10) Hussain, S.M., Brunetti, A., Lucarelli, G., Memeo, R., Bevilacqua, V. and Buongiorno, D., 2022. Deep learning-based image processing for robot assisted surgery: a systematic literature survey. IEEE Access, 10, pp.122627-122657
11) Biswas, P., Sikander, S. and Kulkarni, P., 2023. Recent advances in robot-assisted surgical systems. Biomedical Engineering Advances, p.100109.
12) Zinchenko, K., Wu, C.Y. and Song, K.T., 2016. A study on speech recognition control for a surgical robot. IEEE Transactions on Industrial Informatics, 13(2), pp.607-615.
13) Deshmukh, A.U., Aher, P.D., Kakade, M.S. and Bhise, S.D., Voice Control Robot for Object Detection.
14) van Delden, S., Umrysh, M., Rosario, C. and Hess, G., 2012. Pick-and-place application development using voice and visual commands. Industrial Robot: An International Journal, 39(6), pp.592-600.
15) Ramgire, J.B. and Jagdale, S.M., 2016, August. Speech control pick and place robotic arm with flexiforce sensor. In 2016 International Conference on Inventive Computation Technologies (ICICT) (Vol. 2, pp. 1-5). IEEE.
16) Li, S., Wang, J., Dai, R., Ma, W., Ng, W.Y., Hu, Y. and Li, Z., 2024. RoboNurse-VLA: Robotic Scrub Nurse System based on Vision-Language-Action Model. arXiv preprint arXiv:2409.19590.
17) Kumar, S., Jyothi, A.P., Sumanth, S. and Nagamani, H.S., 2024, March. Recognition and Classification of Apple and Sugarcane Plant Leaf Diseases using SVM with DAE Models. In 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT) (pp. 1-7). IEEE.
18) Pawar, S., Shedge, S., Panigrahi, N., Jyoti, A.P., Thorave, P. and Sayyad, S., 2022, May. Leaf disease detection of multiple plants using deep learning. In 2022 international conference on machine learning, big data, cloud and parallel computing (COM-IT-CON) (Vol. 1, pp. 241-245). IEEE.

19) Jyothi, A.P., Megashree, C., Radhika, S. and Shoba, N., 2021. Detection of cervical cancer and classification using texture analysis. The journal of contemporary issues in business and government, 27(3), pp.1715-1724.