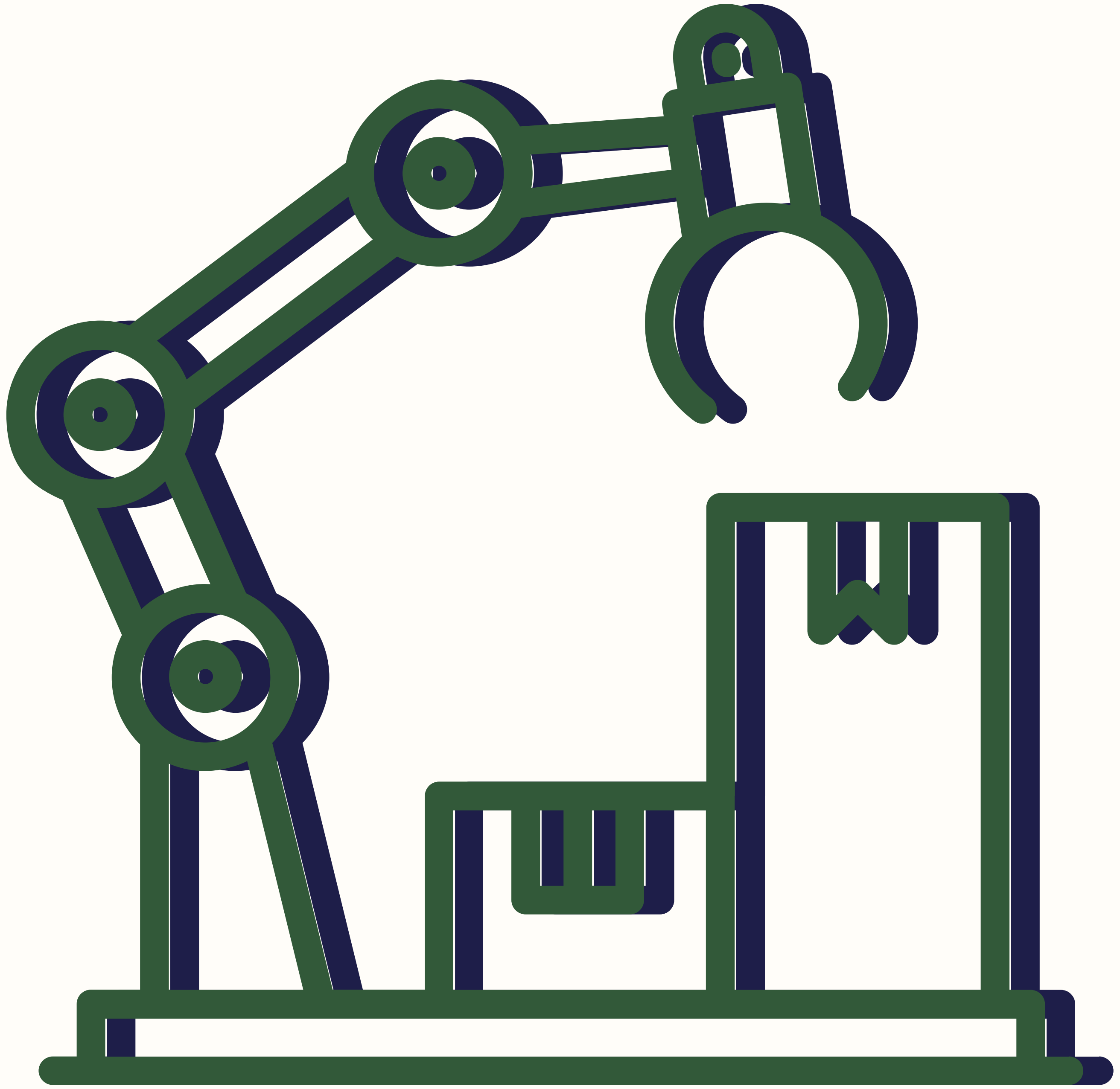
Voice Automated Helping Hand using Object Detection

Literature Review

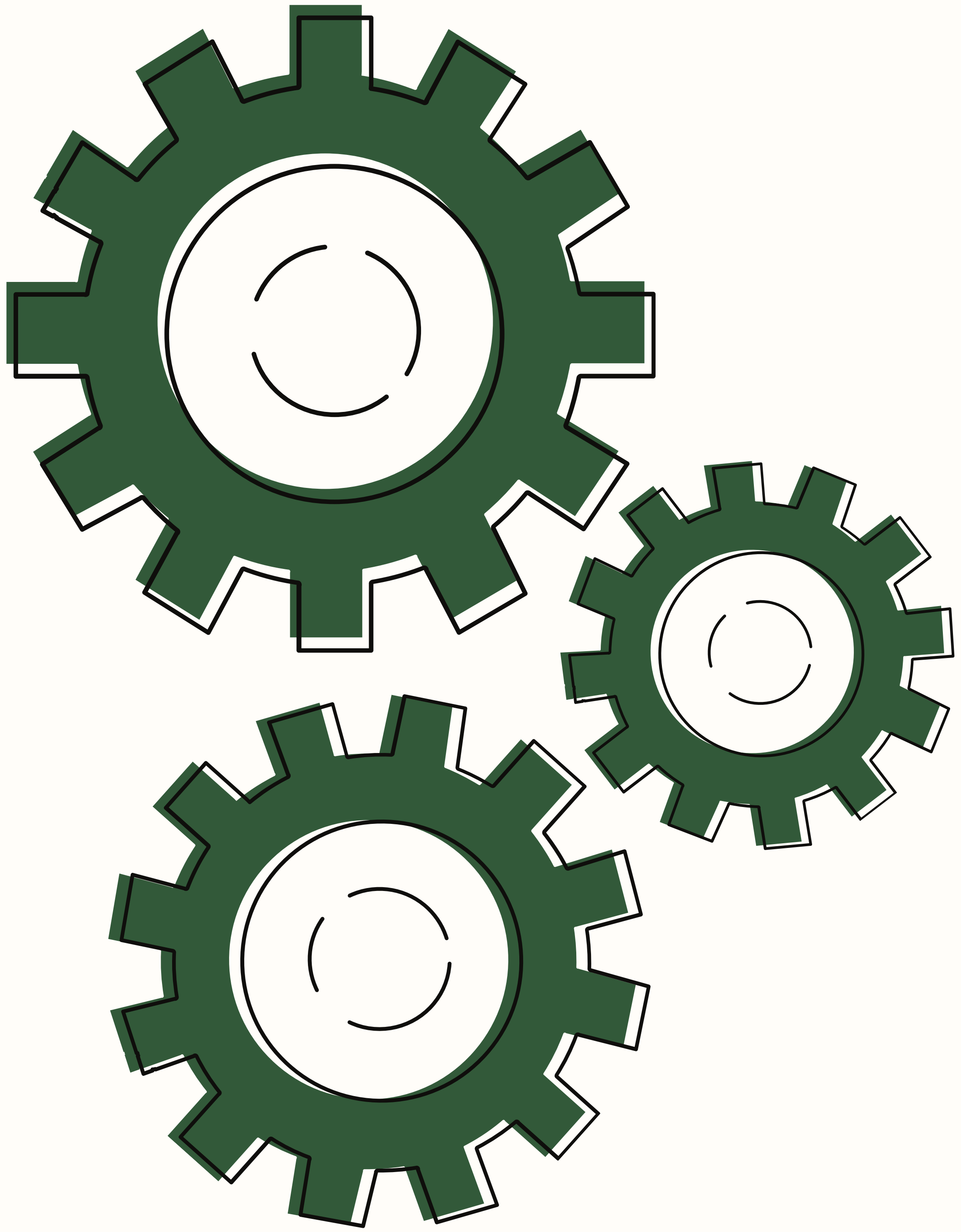
Nigel C. Saldanha (22B-EE-027)
Pritesh Falkar (22B-EE-038)
Ashwesh Mayekar (22B-EE-006)
Atharva Gayakwad (22B-EE-009)



Project Guide: Dr. Sangeeta Mahaddalkar

Agenda

1. Introduction
2. Abstract
3. Literature Review
4. Primary and Secondary Objectives
5. Operations flowchart
6. Block diagram, component specification and B.O.M
7. Estimated timeline
8. References



1. Introduction

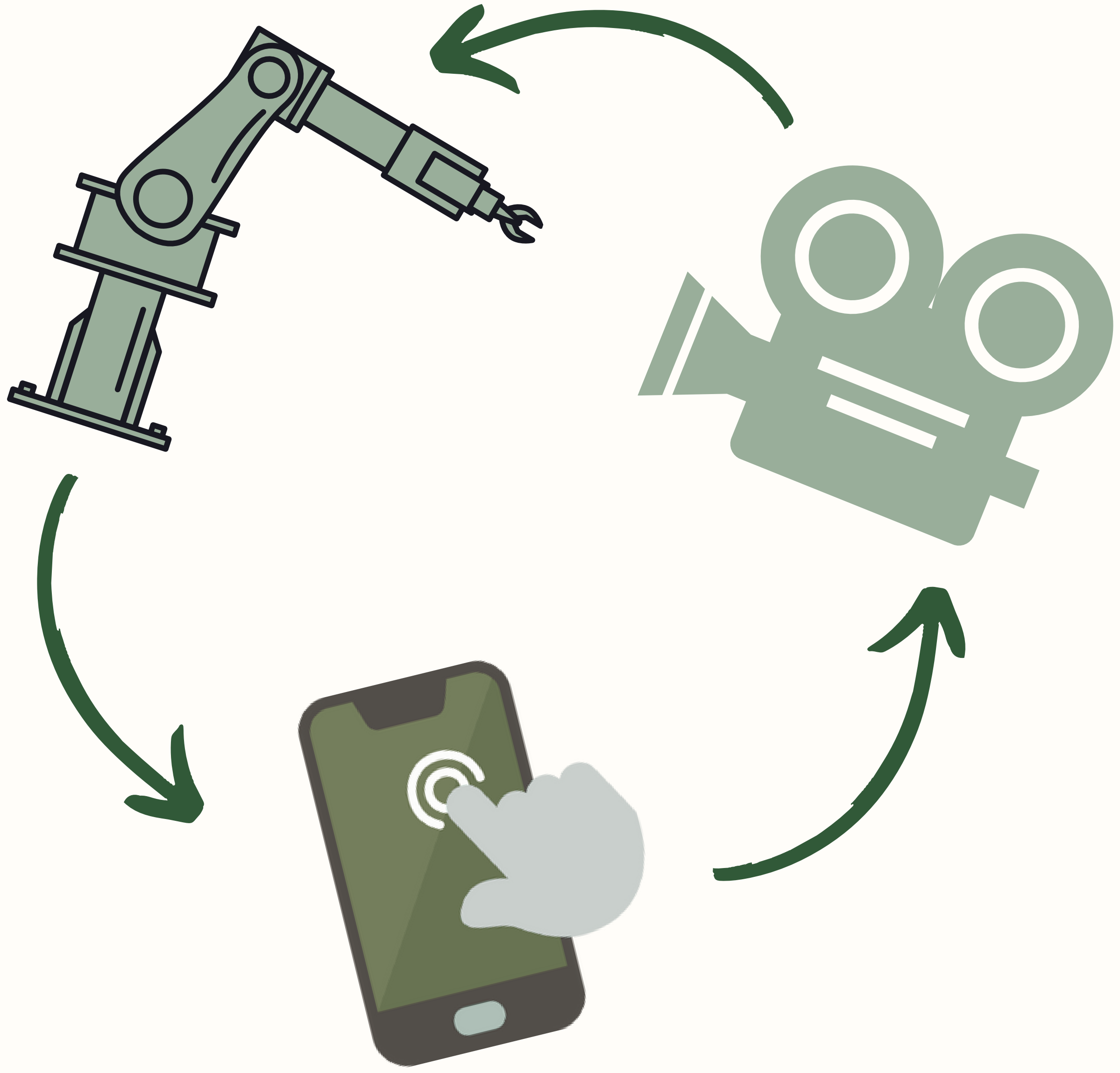
- **Elderly and physically challenged individuals** often face difficulties in performing simple daily tasks like retrieving household objects.
- Existing robotic assistive devices are either too **complex**, **expensive**, or **not user-friendly** for independent use.
- Many available solutions lack natural interfaces (like voice control) and struggle with accurate object detection and reliable grasping.
- There is a gap in designing a **low-cost, intuitive, and accurate** helping hand system that can be practically deployed at home.

Our Approach:

- Develop a **Helping Hand** using:
    - Voice-based commands via a simple mobile app.
    - Dual-camera object detection.
    - A robotic arm modeled with inverse kinematics for reliable grasping.
- Aim: To provide ease of use, high accuracy, and adaptability for assistive living.

2. Abstract

This project presents the design and development of a **Voice Automated Helping Hand**, an intelligent robotic arm that assists users in **retrieving household objects** through **simple voice commands**. A **mobile application** serves as the user interface, enabling real-time **speech-to-text conversion** and **intent recognition**. A **dual-camera system** is employed, with a stationary overhead camera for global object localization and an arm-mounted camera for local refinement. Object detection is powered by **YOLO-based models**, while arm movements are modeled using and inverse kinematics. The system prioritizes **ease of use**, **reliability**, and **accuracy**, aiming to serve as a **practical assistive device** for elderly and physically challenged individuals.



3. Literature Review Summary:

S.No.	Paper Title	Author(s)/Year	Technologies/Methodologies	Key Results/findings	Remarks / Notes
[1]	Voice Controlled personal assistant robot for elderly people	Jishnu U.K.; Indu V.; K.J. Ananthakrishnan; Korada Amith; P Sidharth Reddy; Pramod S. 2020 5th International Conference on Communication and Electronics Systems (ICCES)	• Rasp pi 3B • YOLO • HC05 bluetooth mode • Four wheeled	• FSR - shows inverse relationship between resistance and applied force. i.e. voltage rises as resistance decreases. Can be effectively used to measure gripping force (low force-low voltage and vice versa)	• FSR grasping force
[2]	Assistive device for physically challenged person using voice controlled intelligent robotic arm	Ripcy Anna John; Sneha Varghese; Sneha Thankam Shaji; K.Martin Sagayam 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)	• 1.Voice commands • 4 DOF arm • Joystick • Rail system NEMA17 • Google Assistant	• Phoneme based approach acoustic model (86.5% acc) • SSD mobile net regression model (87.3%) • GoogleNet (78.5%) • ImageNet (87.8%)	• Success based on user acceptability • SNR (signal to noise ratio) 74% accuracy
[3]	Trends in service robots for the disabled and the elderly (ISAC-HERO system)	K. Kawamura; M. Iskarous Proceedings of IEEE/RSJ International Conference Intelligent Robots and Systems (IROS'94) Year: 1994 | Conference Paper | Publisher:IEEE	• Fuzzy command voice interpreter • Macro action builder • Blackboard logic for communication • Task planning - complex tasks broken into sub tasks • PID tuning on each joint, transputer based controller • Macvicar Whelan fuzzy	• Success based on user acceptability (user interface, learning and adaptation)	-
[4]	ROS based control of robot using voice recognition	Rajesh Kannan Megalingam; Racharla Shriya Reddy; Yannam Jahnavi; Manaswini Motheram 2019 Third International Conference on Inventive Systems and Control (ICISC) Year: 2019 | Conference Paper | Publisher: IEEE	• HMM model for speech enhancement • Pocket sphinx (offline) • Rasp pi 3B • ROS • Virtebri algo for best path to decode speech	-	• Uses recognizer and mediator programs

S.No.	Paper Title	Author(s)/Year	Technologies/Methodologies	Key Results/findings	Remarks / Notes
[5]	Small scale robot arm design with pick and place mission based on inverse kinematics	Adnan Rafi Al Tahtawi , Muhammad Agni , Trisiani Dewi Hendrawati	• Simple Kinematics equations for 3 DOF arm	-	-
[6]	Design, construction and control of SCARA prototype with 5 DOF	Delond Angelo Jimenez-Nixon; María Celeste Paredes-Sánchez; Alicia María Reyes-Duke 2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT) Year: 2022 | Conference Paper | Publisher: IEEE	• Denavit-Hartenberg algo and inverse kinematics using • Geometric method • 4 DOF	• Height limitation	• Used V-model for design planning and implementation
[7]	Designing 8 Degrees of Freedom Humanoid robotic arm	Le Bang Duc; Mohd Syaifuddin; Truong Trong Toai; Ngo Huy Tan; Mohd Naufal Saad; Lee Chan Wai 2007 International Conference on Intelligent and Advanced Systems Year: 2007 | Conference Paper | Publisher: IEEE	• 8 DOF robotic hand	-	• Shows different grasping techniques in case we want to make the end effector more versatile in picking different objects.
[8]	Swab-bot - an oral swabbing robotic arm	John Varghese Panicker; Divy Jain; Vibodh H. N; T. Vishnu Yeshwanth; Swetha Ramaiah 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) Year: 2023 | Conference Paper | Publisher: IEEE	• Rasp pi 4B • Adafruit servo hat • Rasp pi cam	• Contains useful kinematics equations for 4 DOF arm	• Arm mounted camera
[9]	Voice conversion based augmentation and a Hybrid CNN-LSTM model for improving speaker independent keyword recognition on limiited datasets	Yeshanew Ale Wubet; Kuang-Yow Lian IEEE Access Year: 2022 | Volume: 10 | Journal Article | Publisher: IEEE	• Speaker independent and detects keywords • ACVAE • Data augmentation (creates new slightly altered sample from original to augment data set) • Pytorch, Keras, Tensorflow	• 96% accuracy	• Speakers from 3 different countries, 12 keywords spoken 10 times
[10]	An MFCC-based Secure Framework for Voice Assistant Systems	Syed Fahad Ahmed;Rabeea Jaffari; Moazzam Jawaid; Syed Saad Ahmed;Shahnawaz Talpur 2022 International Conference on Cyber Warfare and Security (ICCWS) Year: 2022 | Conference Paper | Publisher: IEEE	• 1. Mel-frequency cepstral coefficients based user authenticated security	• 1.Trained on 10 authentic users in 5 conditions: 84% accuracy in normal condition and 59% accuracy in illness condition	• VC arms have limited actions and can take up a lot of electricity

S.No.	Paper Title	Author(s)/Year	Technologies/Methodologies	Key Results/findings	Remarks / Notes
[11]	Comparative study on various architecture of YOLO models used in object recognition	Baranidharan Balakrishnan; Rashmi Chelliah; Madhumitha Venkatesan; Chetan Sah 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) Year: 2022 | Conference Paper | Publisher: IEEE	-	• Darknet architecture gives the maximum accuracy among all the other architectures used. Following those come the Keras architecture and the ImageAI library, where in the ImageAI library, Tiny architecture gave the maximum among the other architectures in that library	-
[12]	Design a Human-robot interaction framework to detect household objects	Sadi Rafsan; Safayet Arefin; A. H. M. Mirza Rashedul Hasan; Mohammed Moshiul Hoque 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV) Year: 2016 | Conference Paper | Publisher: IEEE	• 3 methods to detect objects: size, color, position • Haar, openCV_traincascade, GentleAdaBoost	• 88% accuracy • 11.06% false positive	• Text based interaction
[13]	Object detection using YOLO-V8	B Karthika; M Dharssinee; V Reshma; R Venkatesan; R Sujarani 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) Year: 2024 | Conference Paper | Publisher: IEEE	• DETR • EfficientDet • Roboflow for annotating	• 300 epochs, 85% precision, 80% recall, 82% mAP	-
[14]	Design and implementation of a robotic arm assistant with voice interaction using machine vision	Nantzios, G.; Baras, N.; Dasygenis, M. Design and Implementation of a Robotic Arm Assistant with Voice Interaction Using Machine Vision. Automation 2021, 2, 238–251. https://doi.org/10.3390/automation2040015	Nantzios, G.; Baras, N.; Dasygenis, M. Design and Implementation of a Robotic Arm Assistant with Voice Interaction Using Machine Vision. Automation 2021, 2, 238–251. https://doi.org/10.3390/automation2040015	• 90% accuracy (96% with QR) • 90% voice recognition (96% with QR) • Drop item failure 3% • Avg execution time 52secs	• Arm range 200 degrees • Movement speed reduced to 30% of attainable speed
[15]	Robotic arm vehicle using voice recognition for challenged people	Jia Nannda; Lokareddy Venkanna Dora; Tanvi Gupta; Aryan Chouhan; Anmol Verma 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)	• Vehicle • Predefined commands for voice recognition • Ultrasonic sensors	• Total time completion formulas	-

4.1 **Primary objectives:**

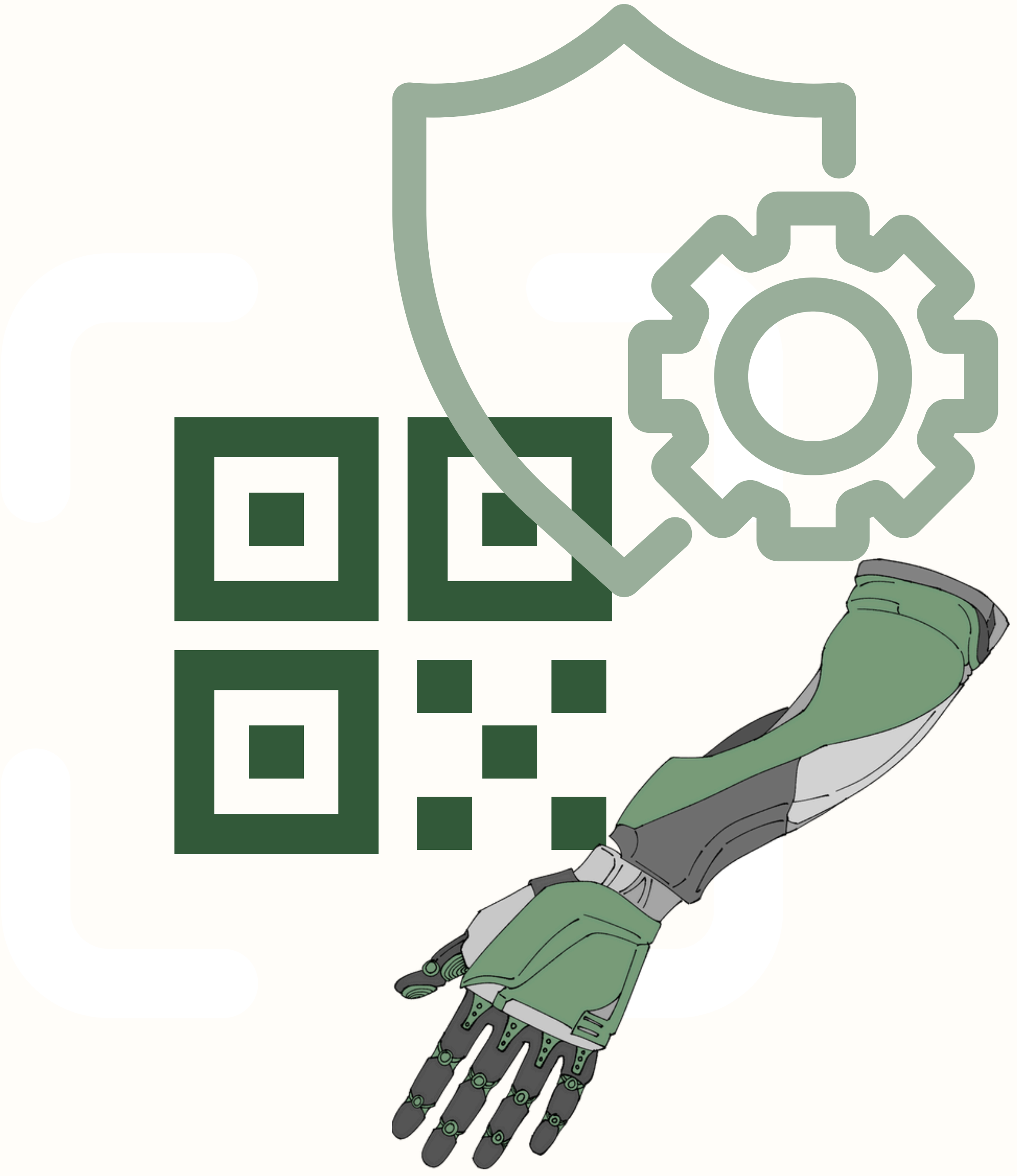
- **Build simple one button app**
  Aim: To design a user friendly app catered to an elderly audience and achieve voice accuracy >90%
  - App using MIT app inventor/ Flutter
  - OpenAI whisper/Vosk/SpeechRecognition for Speech-to-text
  - Tensorflow for intent classifier

- **Using YOLOe/YOLOv12 for object detection**
  Aim: Reduce false positives <11.06% and achieve detection accuracy >90%
  - Dual-camera localization (stationary Darknet based model + dynamic arm mounted TinyYOLO based model)
  - Roboflow for annotating and labelling

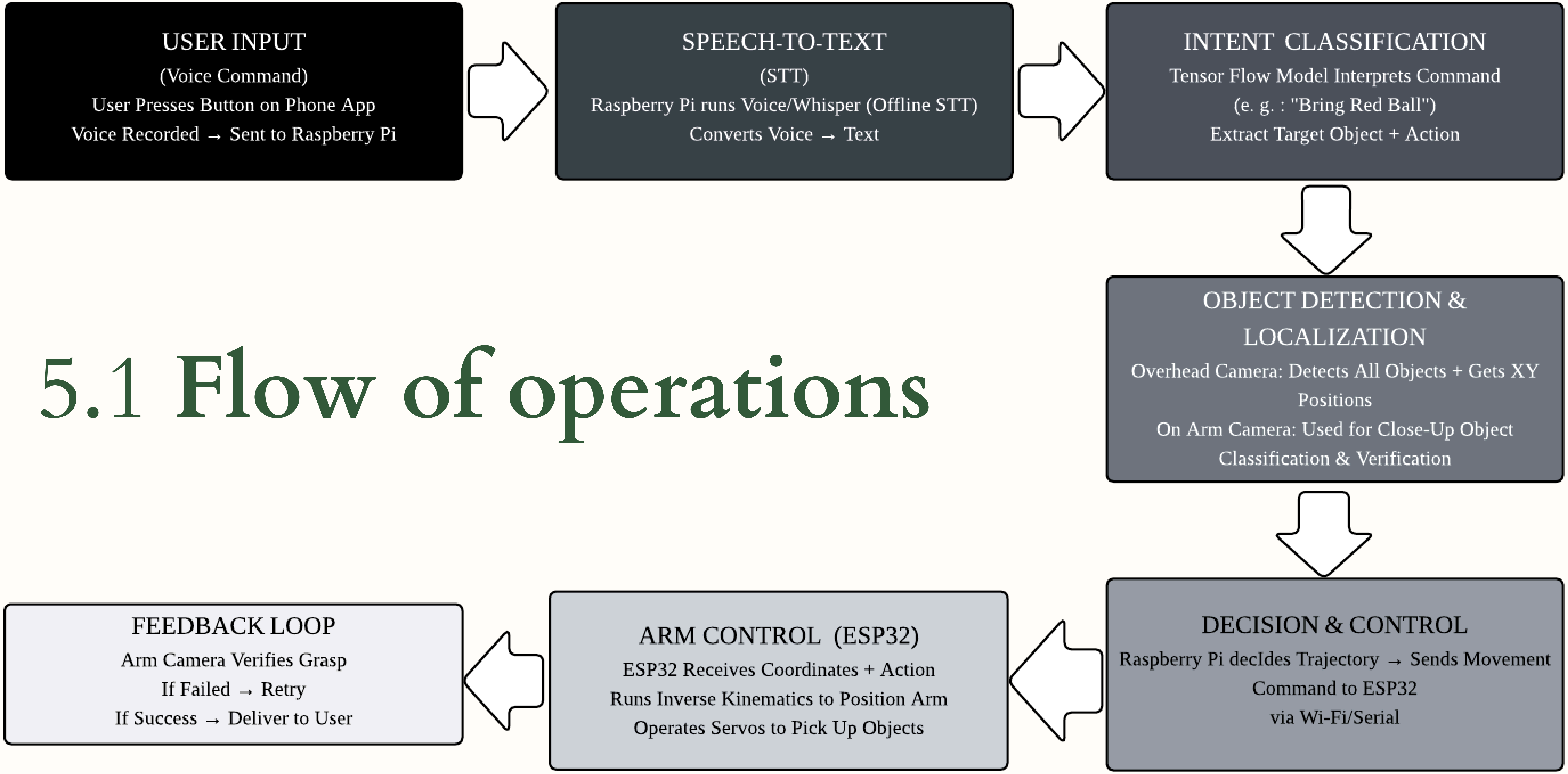
- **Arm mechanics**
  Aim: Reduce drop item failure <3%
  - Denavit-Hartenberg and geometric inverse kinematics to model arm movements mathematically
  - Implement appropriate gripping using FSR

4.2 Secondary objectives:

- Implement rail-based system across workspace to extend horizontal picking range
- Improve upon 74% SNR using HMM + Viterbi decoding
- Fine tune arm movements using PID and picking accuracy
- Lightweight data augmentation to artificially expand small speech datasets
- Use QR code object detection approach for commonly used household items
- Improve upon existing end-effector design to optimize picking objects of different shapes and sizes
- Security layer for voice recognition



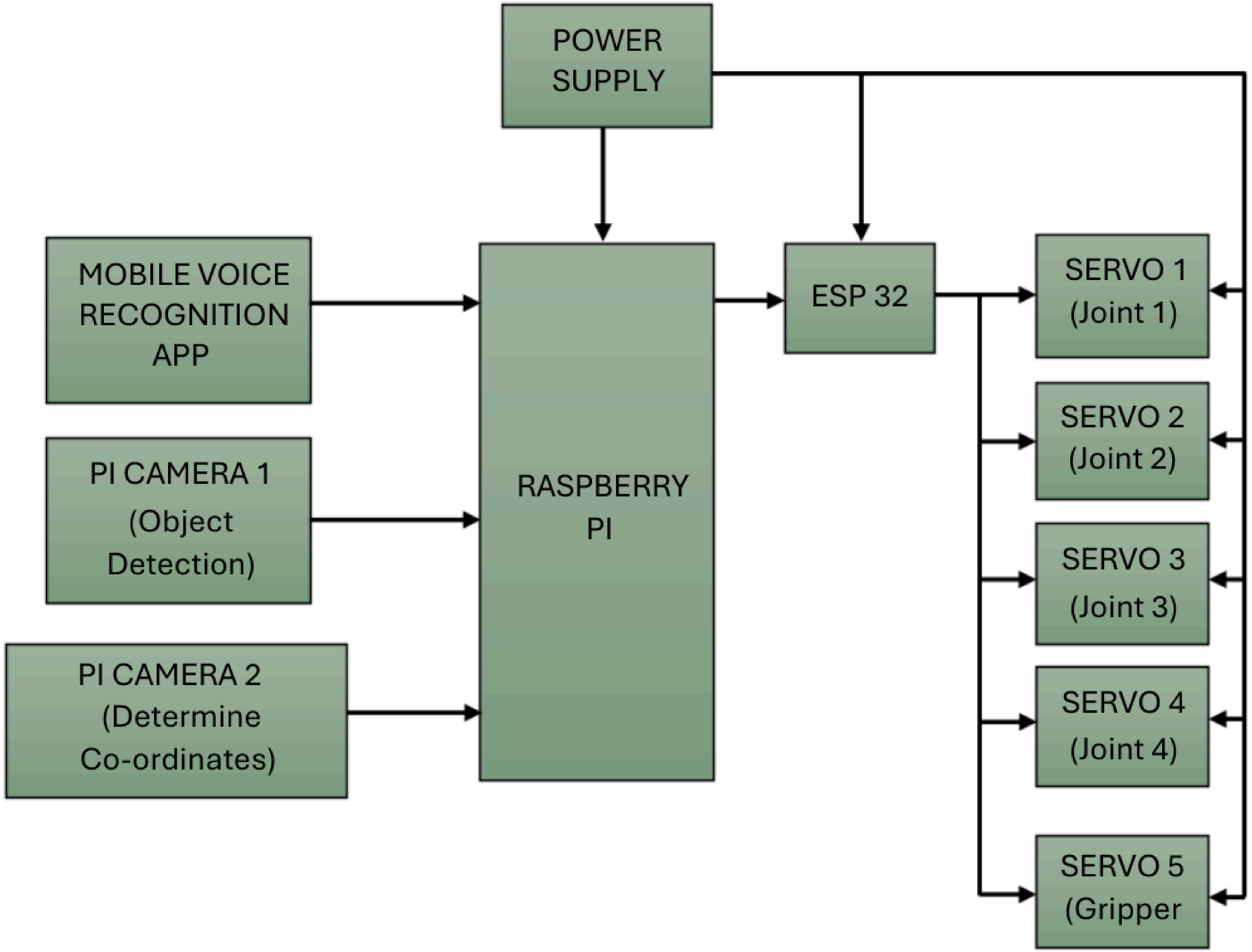
5.1 Flow of operations



5.2 High level overview

- **User Input (App Interface)**
  - The user presses and holds a large button on the app.
  - While holding, the app records the audio command.
  - When released, recording stops and the audio is sent to the bot.
  - The app then converts speech to text and displays it on screen for user verification.
- **Command Verification**
  - If the text matches what the user intended, the app sends it forward.
  - If not, the user can re-record.
  - Once verified, the command is passed to the bot (Raspberry Pi).
- **Intent Classification**
  - The Raspberry Pi runs a TensorFlow-based intent classifier.
  - It interprets the command (e.g., “bring the phone,” “pick up the bottle”).
  - The classifier decides what object needs to be fetched.
- **Stationary Camera and Dynamic Arm-mounted camera**
  - A fixed overhead camera scans the workspace.
  - A YOLO model identifies all visible objects and determines their approximate (x,y) coordinates and the number of objects using primary camera and identification of the objects is done by the dynamic cam.
- **Arm Movement Initialization**
  - The ESP32 controlling the robotic arm receives the target coordinates.
  - Using Denavit-Hartenberg + inverse kinematics, the arm is moved toward the rough position.
- **Grasp Execution**
  - The end-effector (gripper) positions itself grasps the object.
- **Object Retrieval**
  - Once the object is secured, the arm lifts it.
  - The system calculates the path back (simple return path).

6.1 Block Diagram



6.2 Component specifications

1) Raspberry Pi 4 Model B

Processor	Broadcom BCM2711, Quad-core Cortex-A72 @ 1.8 GHz
Memory	4 GB
Connectivity	Dual-band Wi-Fi, Bluetooth 5.0, Gigabit Ethernet, 2× USB 3.0, 2× USB 2.0
Video/Audio	2× micro-HDMI (4Kp60), MIPI DSI/CSI, audio out
Storage	Micro-SD card slot
Power	5V DC via USB-C (≥3A), PoE supported
GPIO	40-pin header
Temperature	0–50°C

2) ESP 32-WROOM

Processor	Dual-core 32-bit LX6, up to 240 MHz
Memory	520 KB SRAM, 4–16 MB flash support
Connectivity	Wi-Fi 802.11 b/g/n, Bluetooth 4.2 (Classic + BLE)
GPIO Pins	34 (PWM, ADC, DAC, UART, I²C, SPI)
PWM/ADC/DAC	16 PWM channels, 12-bit ADC, 2× 8-bit DAC
Voltage	3.3 V typical
Current	~160 mA (Wi-Fi active), <10 µA (deep sleep)
Size	18 × 25.5 mm

3) MG996R Servo Motor

Type	High-torque metal-geared servo
Torque	11 kg-cm @ 6V
Rotation Angle	~360°
Voltage	4.8 – 7.2 V

4) MG90S Micro Servo

Type	Metal-geared micro servo
Torque	2.2 kg-cm @ 6V
Rotation	180°
Voltage	4.8 – 6 V

5) Raspberry Pi Camera Module 3

Sensor	Sony IMX708, 12.3 MP
Focus	Phase Detection Autofocus (PDAF), 10 cm – ∞
Video	1080p@50fps, 720p@100fps, 480p@120fps
Field of View	75° (Standard)
Variants	Standard, Wide, NoIR (for night vision)
Interface	CSI connector (200 mm ribbon)
Size	25 × 24 mm (approx.)

6.3 **Estimated B.O.M.**

Sr. No	Components	Quantity	Estimated Price
1	Raspberry PI 4 Model B	1	5,758 Rs
2	ESP 32-WROOM	1	499 Rs
3	MG996R Servo Motor	3	1,119 Rs
4	MG90S Micro Servo	2	378 Rs
5	Raspberry PI Camera Module 3	2	6,198 Rs
		Total Price	13,952 Rs

7. Project Timeline

Week 1-2
Oct 1 – Oct 14
STT Method

Week 3-4
Oct 15 – Oct 28
Tensorflow

Week 5-6
Oct 29 – Nov 11
YOLO

Week 7-8
Nov 12 – Nov 25
Prototype App

Phase 1: Software implementation (ETC: December)

Phase 2: Arm implementation and calibration (ETC: April)

Week 9-12
Nov 26 – Dec 23
Simulation

Week 13-16
Jan 2 – Jan 29
Arm prototype

Week 17-20
Jan 30 – Feb 26
Implementation

Week 21-24
Feb 27 – Mar 26
Revisions

8. References

- **[1]** J. U.K., I. V., K. J. Ananthakrishnan, K. Amith, P. S. Reddy and P. S., Voice Controlled Personal Assistant Robot for Elderly People, 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 269-274, doi: 10.1109/ICCES48766.2020.9138101.
- **[2]** R. A. John, S. Varghese, S. T. Shaji and K. M. Sagayam, Assistive Device for Physically Challenged Persons Using Voice Controlled Intelligent Robotic Arm, 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 806-810, doi: 10.1109/ICACCS48705.2020.9074236.
- **[3]** K. Kawamura and M. Iskarous, Trends in service robots for the disabled and the elderly, Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'94), Munich, Germany, 1994, pp. 1647-1654 vol.3, doi: 10.1109/IROS.1994.407636.
- **[4]** R. K. Megalingam, R. S. Reddy, Y. Jahnavi and M. Motheram, ROS Based Control of Robot Using Voice Recognition, 2019 Third International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2019, pp. 501-507, doi: 10.1109/ICISC44355.2019.9036443.
- **[5]** Al Tahtawi, Adnan & Agni, Muhammad & Hendrawati, Trisiani. (2021). Small-scale Robot Arm Design with Pick and Place Mission Based on Inverse Kinematics. Journal of Robotics and Control (JRC). 2. 10.18196/26124.
- **[6]** D. A. Jimenez-Nixon, M. C. Paredes-Sánchez and A. M. Reyes-Duke, Design, construction and control of a SCARA robot prototype with 5 DOF, 2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), Soyapango, El Salvador, 2022, pp. 1-6, doi: 10.1109/ICMLANT56191.2022.9996479.
- **[7]** L. B. Duc, M. Syaifuddin, T. T. Toai, N. H. Tan, M. N. Saad and L. C. Wai, Designing 8 Degrees of Freedom Humanoid Robotic Arm, 2007 International Conference on Intelligent and Advanced Systems, Kuala Lumpur, Malaysia, 2007, pp. 1069-1074, doi: 10.1109/ICIAS.2007.4658549.

- **[8]** Panicker, John & Jain, Divy & N, Vibodh & Yeshwanth, T. & Ramaiah, Swetha. (2023). SwabBot-An oral Swabbing Robotic arm. 1-6. 10.1109/ICECCME57830.2023.10252564.
- **[9]** Y. A. Wubet and K. -Y. Lian, Voice Conversion Based Augmentation and a Hybrid CNN-LSTM Model for Improving Speaker-Independent Keyword Recognition on Limited Datasets, in IEEE Access, vol. 10, pp. 89170-89180, 2022, doi: 10.1109/ACCESS.2022.3200479.
- **[10]** S. F. Ahmed, R. Jaffari, M. Jawaid, S. S. Ahmed and S. Talpur, An MFCC-based Secure Framework for Voice Assistant Systems, 2022 International Conference on Cyber Warfare and Security (ICCWS), Islamabad, Pakistan, 2022, pp. 57-61, doi: 10.1109/ICCWS56285.2022.9998446.
- **[11]** B. Balakrishnan, R. Chelliah, M. Venkatesan and C. Sah, Comparative Study On Various Architectures Of Yolo Models Used In Object Recognition, 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 685-690, doi: 10.1109/ICCCIS56430.2022.10037635.
- **[12]** S. Rafsan, S. Arefin, A. H. M. M. R. Hasan and M. M. Hoque, Design a human-robot interaction framework to detect household objects, 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, 2016, pp. 973-978, doi: 10.1109/ICIEV.2016.7760144.
- **[13]** B. Karthika, M. Dharssinee, V. Reshma, R. Venkatesan and R. Sujarani, Object Detection Using YOLO-V8, 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-4, doi: 10.1109/ICCCNT61001.2024.10724411.
- **[14]** Nantzios, G.; Baras, N.; Dasygenis, M. Design and Implementation of a Robotic Arm Assistant with Voice Interaction Using Machine Vision. Automation 2021, 2, 238-251.
- **[15]** J. Nannda, L. V. Dora, T. Gupta, A. Chouhan and A. Verma, A Robotic Arm Vehicle using Voice Recognition for Physically Challenged People, 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), Faridabad, India, 2024, pp. 858-862, doi: 10.1109/ICAICCIT64383.2024.10912202.

Thank you!