International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 11 | Nov - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Human Following Robot for Elderly Care with Medication Reminder and Emotion Recognition

Bhavana S[1], Girish C M[2], Kavya Vaggar[3], Lisha T[4], Rahul R Rai[5]

[1-4]Students, Dept. of Electronics and Communication, SJB Institute of Technology, Bengaluru, Karnataka, India
[5]Assistant Professor, Dept. of Electronics and Communication, SJB Institute of Technology, Bengaluru, Karnataka, India

---

Abstract - Elderly people often need steady help with getting around, taking their medicines on time and looking after their emotional well-being, and much of that load falls on caregivers. This paper describes a low-cost assistive robot that follows a person around and brings several supports together on one mobile platform: person tracking, obstacle avoidance, medication reminders, emotion recognition from voice, face recognition and caregiver alerts. A Raspberry Pi 4 acts as the main controller. A Pi Camera feeds an OpenCV and YOLO person-detection pipeline, and its output steers the DC motors through a motor driver, while an ultrasonic sensor keeps the robot away from obstacles. A Real-Time Clock (RTC) scheduler checks the stored medication timetable against the current time and raises alerts on an LCD, a buzzer or a speaker, with optional caregiver notifications sent through the Telegram Bot API. Speech picked up by a microphone is turned into pitch, energy and MFCC features, which a trained CNN or LSTM model sorts into emotion categories. In our prototype, scheduled reminders, caregiver notifications, voice emotion prediction and face recognition all work together. The robot is meant to help caregivers, not to take the place of professional medical care or human supervision.

Key Words: **Assistive robot, elderly care, human following, YOLO, medication reminder, speech emotion recognition.**

1. INTRODUCTION

The world's elderly population keeps growing, and many older adults live alone or go for long stretches with no caregiver nearby. Everyday tasks get harder with age: moving around the house, remembering to take medicines at the right times, and making emotional needs known. Poor adherence to prescribed medication is a well-recognised problem [1], and when caregivers are away, loneliness or low mood can easily go unnoticed [2]. Robots suit this kind of support well, because they can move with the user, watch their surroundings and communicate through sound and a display [2]. Still, most simple solutions cover just one need. A pill-box alarm cannot follow the user, a follower robot does nothing for their health, and a voice assistant cannot go from room to room with them.

In this paper we propose a human-following assistive robot for elderly care, with medication reminders and emotion recognition. A Raspberry Pi 4 runs the robot. An OpenCV and YOLO pipeline finds the person, an ultrasonic sensor spots obstacles, and a Real-Time Clock (RTC) scheduler triggers medication reminders on an LCD and a buzzer or speaker. A microphone-based module classifies the emotion in the user's voice, and selected alerts reach caregivers through the Telegram Bot API over Wi-Fi. The robot is there to assist caregivers; it does not replace professional medical care or human supervision. Our main contributions are:

- A low-cost robot that combines human following, obstacle avoidance, medication reminders, voice emotion recognition, face recognition and caregiver notification on a single Raspberry Pi 4.
- A camera-based following method in which the horizontal centre of the YOLO bounding box decides whether the robot steers left, straight or right, with an ultrasonic check before every movement.
- An RTC-based reminder module that does not fire the same alert twice, and a voice emotion module that uses pitch, energy and MFCC features with a CNN or LSTM classifier.

2. RELATED WORK

Region-proposal detectors such as R-CNN [3] are accurate but too heavy for most embedded boards. Single-shot detectors such as SSD [4] and YOLO [5] predict boxes and class scores in one pass through the network, which makes them a better fit for real-time use. OpenCV [6] handles the frame capture and pre-processing that connect these detectors to low-cost cameras, and ultrasonic range sensors are commonly paired with them so a robot can stop or turn aside near obstacles. Assistive and social robots for elderly care are reviewed in [2]. Among their functions, medication reminders are some of the most practical, since forgetting doses is a known barrier to adherence [1]. Pairing reminders with caregiver notifications through a messaging service such as the Telegram Bot API [7] gives caregivers remote awareness without any dedicated infrastructure.

Speech emotion recognition works out emotional categories from acoustic cues, and [8] surveys the features, classifiers and databases involved. Mel-frequency cepstral coefficients (MFCCs) [9], along with pitch and energy, are the most widely used features. Deep models [10], including CNNs and long short-term memory (LSTM) networks [11], learn patterns from these features and are often trained on acted corpora such as RAVDESS [12]. Such classifiers may not carry over well to elderly speakers or to noisy homes. Most earlier work deals with detection, reminders or emotion recognition on their own; we bring all three together on one low-cost mobile platform.

3. PROPOSED SYSTEM DESIGN

3.1 System Architecture

Fig. 1 shows the architecture, which has four parts. **Input devices:** The Pi Camera captures video for person detection, the ultrasonic sensor measures the distance to obstacles, the microphone records the user's voice, and the RTC keeps time for medication scheduling. **Processing unit:** The Raspberry Pi 4 runs Python, OpenCV and YOLO for detection, along with the audio processing and scheduling logic. **Decision and control modules:** These modules handle following, obstacle detection, reminders and emotion recognition, and they use the person's position and any nearby obstacles to decide how the robot moves. **Output and communication devices:** A motor driver and DC motors move the robot, an LCD and a buzzer or speaker give alerts, and Wi-Fi with the Telegram Bot API notifies caregivers. Because the design is modular, each function can be developed and tested on its own.



Fig -1: System architecture of the proposed robot.

3.2 Hardware and Software Components

On the hardware side, the robot uses a Raspberry Pi 4 as the central controller, a Pi Camera for live video, an ultrasonic sensor for obstacle distance, and a motor driver with DC motors for steering and motion. A 16×2 LCD and a buzzer show reminders and system messages, and a speaker can be added for voice announcements. An RTC module handles schedule timing and a microphone takes voice input. The software is written in Python. OpenCV handles image processing and YOLO detects people, while GPIO libraries talk to the hardware. Audio-processing libraries and a trained classification model take care of emotion recognition, scheduling logic drives the reminders, and the Telegram Bot API carries messages to caregivers.

3.3 Working Principle

When the robot powers up, the Raspberry Pi initialises all devices and software modules. The camera streams continuously and YOLO looks for people in every frame. Depending on where the person is, the controller sends movement commands to the motors. Meanwhile the ultrasonic sensor scans for obstacles, and if one comes within a set safety distance the robot either halts or takes a safe detour. The reminder module compares the RTC time with the stored schedule, and when a reminder falls due it lights the LCD and sounds the buzzer or speaker. Microphone input is processed and classified, and the predicted emotion can be displayed or logged. When Wi-Fi is available, selected alerts go to caregivers through Telegram.

4. HUMAN FOLLOWING AND NAVIGATION

The Pi Camera streams frames to the Raspberry Pi 4. There, OpenCV pre-processes each frame and YOLO detects people and draws a bounding box around them. The horizontal centre of the box is

$$x^c = (x_{min} + x_{max}) / 2 \quad (1)$$

where x_{min} and x_{max} are the left and right edges of the box. We split the frame into left, centre and right zones. If the person is on the left, the robot turns left; if they are on the right, it turns right; and when they are near the centre, it moves forward only if the path is clear and the distance is safe. If no person is found, the robot stops or pauses and searches again. A dead zone in the centre prevents needless turning when the person shifts only slightly, and the motor driver turns the robot by changing the direction or speed of the left and right motors.

For obstacle avoidance, the ultrasonic sensor sends out a sound pulse and times how long the echo takes to return, so the distance is

$$d = (v \times t) / 2 \quad (2)$$

where v is the speed of sound in air and t is the round-trip time. We divide by two because the pulse travels to the obstacle and back. The controller compares d with a safety threshold. If the path is clear, the robot keeps following; if not, it stops or detours and checks its surroundings again before moving on. The algorithm runs in a continuous loop: (i) initialise the system, (ii) capture a frame, (iii) detect the person, (iv) decide whether the box centre is left, centre or right, (v) check for obstacles, (vi) turn or move forward if it is safe, (vii) stop or pause if detection fails, and (viii) start over.

5. MEDICATION REMINDER MODULE

Taking medicines on time matters a great deal for elderly users, who often have several to take at set times. This module brings together the RTC, the Raspberry Pi 4, the LCD and the buzzer. A schedule listing each medicine name and its reminder time is stored in the program or in a schedule file. The RTC supplies the current time, and with a backup battery it keeps the time even when the main system is off. The controller compares the current time with the stored entries, and when a scheduled time arrives it sounds the buzzer and shows a message such as “MEDICINE TIME – Tablet: [Medicine Name]” on the LCD. The medicine name here is only an example; real entries should follow the user's prescription. A record of reminders that have already fired stops the same alert from repeating throughout the matching time interval.

If Wi-Fi and the Telegram bot are set up, the caregiver also gets a notification with the reminder event, its time and the robot's status. The procedure runs as follows: initialise the RTC, LCD and buzzer; load the schedule; read the time; compare it with the schedule; generate the alert; notify the caregiver; record the event; and keep monitoring. The module gives visual and audible alerts and lets caregivers keep an eye on things remotely, but it has limits. Wrong schedule entries, a power cut, an incorrectly set RTC or a lost internet connection all affect how it works, and since there is no way for the user to acknowledge a reminder, the system cannot confirm that the medicine was actually taken. It should therefore be seen as a supportive reminder, not a substitute for medical advice or caregiver supervision.

6. EMOTION RECOGNITION MODULE

The module records speech through the microphone and prepares it by segmenting it, reducing noise, normalising amplitude, framing and windowing it, and checking that the audio is long and clear enough. It then extracts three acoustic features. *Pitch* is related to the fundamental frequency of voiced speech, though it varies between individuals and speaking styles. *Energy* reflects the intensity of the signal, but microphone distance and recording volume affect it, so a loud voice does not necessarily mean anger. *MFCCs* give a compact description of the short-term speech spectrum. To get them, the signal is framed and windowed, its spectrum is passed through a Mel-scaled filter bank, logarithms are taken, and a Discrete Cosine Transform (DCT) is applied.

The features go into a trained CNN or LSTM classifier. A CNN picks up local patterns in the feature representation, while an LSTM, a recurrent network built to handle temporal dependencies, looks at how the features change over the course of an utterance. The model scores the supported classes (happy, sad, angry, fear and neutral) and picks the most likely one. These outputs are predictions from a model, not confirmed emotional states. One utterance may not reflect a person's overall state, so a prediction should never be treated as an emergency or used as a basis for medical decisions. Noise,

accent, language and the training data all affect performance, and a model trained on acted speech may not work as well for elderly users, so it needs to be evaluated using accuracy, precision, recall and a confusion matrix. Voice data is also sensitive, so it should be collected sparingly and protected.

7. IMPLEMENTATION AND RESULTS

The Raspberry Pi manages the overall operation. Camera video is processed with OpenCV and YOLO to identify and track the person, and the detected position decides whether the robot moves forward, turns left, turns right or stops, through the motor driver and DC motors. The ultrasonic sensor helps it avoid collisions. A face recognition module identifies the person in front of the robot and reports their name, and a rechargeable battery powers the electronics. Fig. 2 shows the prototype.



Fig -2: Hardware prototype of the robot.

Figs. 3–6 show outputs from the prototype. Fig. 3 is the reminder interface, where the medicine name, date and time are entered and listed in a schedule table; a pop-up announces the reminder at the scheduled time, and a message confirms when a reminder has been added. Fig. 4 shows the 16×2 LCD displaying a reminder along with the medicine name. Fig. 5 is the voice emotion interface, which accepts an uploaded audio file or a five-second recording and shows the predicted emotion with a confidence value (Happy in this example). Fig. 6(a) shows the reminders a caregiver receives through the Telegram bot, and Fig. 6(b) shows the face recognition output, where a face is detected, encoded, matched against a stored image and named.



Fig -3: Medicine reminder interface: (a) reminder pop-up and (b) reminder added.



Fig -4: Reminder displayed on the LCD.



Fig -5: Voice emotion detection interface.



Fig -6: (a) Telegram caregiver notification and (b) face recognition output.

The prototype demonstrates the main strengths of the system: camera-based following, timely medication reminders, voice emotion detection, face recognition and Telegram notifications to the caregiver, all on affordable, widely available components that suit academic development. Performance may depend on lighting, camera position, obstacles and processing speed, and emotion accuracy may vary with the speaker, background noise and recording quality. More testing is needed before the robot could be trusted to operate independently in real elderly-care settings.

8. CONCLUSION AND FUTURE WORK

The proposed robot combines a Raspberry Pi 4, computer vision, speech-based emotion recognition and embedded hardware. It follows a person, gives timely medication reminders, recognises emotions from speech, identifies faces and keeps caregivers informed. In doing so it supports the daily activities of elderly people and reduces the need for continuous manual support. Future work includes voice-command interaction, GPS tracking and emergency notification; more advanced, noise-robust emotion models trained on representative data; cloud storage of medication schedules and caregiver updates; and better obstacle detection, navigation and battery backup.

REFERENCES

1. World Health Organization, Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: WHO, 2003.
2. J. Broekens, M. Heerink, and H. Rosendal, “Assistive social robots in elderly care: A review,” Gerontechnology, vol. 8, no. 2, pp. 94–103, 2009.
3. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proc. IEEE CVPR, 2014, pp. 580–587.
4. W. Liu et al., “SSD: Single shot multibox detector,” in Proc. ECCV, 2016, pp. 21–37.
5. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE CVPR, 2016, pp. 779–788.
6. G. Bradski, “The OpenCV library,” Dr. Dobb's Journal of Software Tools, 2000.
7. Telegram, “Telegram Bot API.” [Online]. Available: https://core.telegram.org/bots/api
8. M. El Ayadi, M. S. Kamel, and F. Karray, “Survey on speech emotion recognition: Features, classification schemes, and databases,” Pattern Recognition, vol. 44, no. 3, pp. 572–587, 2011.
9. S. B. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” IEEE Trans. Acoust.,

Speech, Signal Process., vol. 28, no. 4, pp. 357–366, Aug. 1980.
10. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.
11. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
12. S. R. Livingstone and F. A. Russo, “The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS),” PLoS ONE, vol. 13, no. 5, e0196391, 2018.