Research Article

AI-Powered Voice Recognition and Autonomous Navigation for Mobile Robots

[1]M. Sasikala, [2]A. Divya, [3]K. Tamilselvi, [4]Manmohan Singh, [5]A.Saranya, [6]L.Sarojini

[1]Assistant Professor, Department of Electrical and Electronics Engineering, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Trichy, Tamilnadu, India. msasikala06.05.1992@gmail.com
[2]Assistant Professor, Department of Electrical and Electronics Engineering, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Trichy, Tamilnadu, India divyaa.jc@gmail.com
[3]Assistant Professor, Department of Electrical and Electronics Engineering, NPR College of Engineering & Technology,Natham,Dindigul District,Tamilnadu. tamilselvi@nprcolleges.org
[4]Assistant professor, Department of computer science and engineering, Galgotias University, Greater Noida, Uttar Pradesh,India. manmohan.singh@galgotiasuniversity.edu.in
[5]Assistant Professor, Department of Electrical and Electronics Engineering, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Trichy, Tamilnadu, India. asaranyaeee2012@gmail.com
[6]Assistant Professor, Department of Electrical and Electronics Engineering, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Trichy, Tamilnadu, India. sarojinil.set@dsuniversity.ac.in

Abstract

The integration of artificial intelligence (AI) with autonomous robotic systems has enabled more natural and intelligent human–robot interaction. This study presents an AI-driven voice recognition and autonomous control system for mobile robots, designed to interpret spoken commands and execute corresponding robotic actions in real time. A machine learning-based voice recognition approach is employed to classify predefined speech commands and convert them into control instructions. The recognized commands are integrated with autonomous navigation and obstacle detection modules to enable the robot to move safely toward the desired direction while responding to human instructions. The proposed system evaluates voice recognition accuracy, response time, navigation accuracy, and obstacle detection performance under different operating conditions. Experimental results demonstrate a voice recognition accuracy of 99.2%, with a response time of approximately 200 ms. The system also achieves 98.8% navigation accuracy and 98.9% obstacle detection accuracy, demonstrating effective integration of AI-based speech processing with autonomous robotic control. However, noisy environmental conditions reduce recognition performance, indicating the need for improved noise-robust training and preprocessing techniques. The proposed approach provides an intelligent and efficient framework for voice-controlled autonomous mobile robots and has potential applications in smart automation, industrial environments, assistive robotics, and human–robot interaction.

Keywords: Artificial Intelligence, Machine Learning, Voice Recognition, Speech Recognition, Autonomous Robot, Mobile Robot, Autonomous Navigation, Obstacle Detection.

1. Introduction

Artificial intelligence has become an important technology for developing autonomous robotic systems capable of interacting with humans and making intelligent decisions. Conventional robot control systems generally depend on physical switches, remote controllers, or predefined input mechanisms, which can limit flexibility and natural interaction. Voice-based control provides a more intuitive interface by allowing users to communicate with robots through spoken commands [1]-[3]. The combination of speech recognition, machine learning, autonomous navigation, and obstacle detection can therefore improve the intelligence and usability of mobile robotic systems.

Recent developments in machine learning and deep learning have significantly improved the ability of computer systems to recognize and classify human speech [4]. Machine learning algorithms can extract relevant features from voice signals and identify predefined commands with high accuracy. However, voice recognition in real-world environments remains challenging because of background noise, variations in pronunciation, speaker characteristics, and changes in environmental conditions[5]-[8]. Therefore, an effective robotic voice-control system requires appropriate preprocessing, feature extraction, classification, and decision-making mechanisms.

Autonomous navigation is another essential component of intelligent mobile robots. After receiving a voice command such as *forward, backward, left,* or *right*, the robot must translate the recognized command into an appropriate movement while continuously monitoring its surroundings [9]. Obstacle detection sensors can identify objects in the robot's path and prevent collisions. Thus, integrating AI-based voice recognition with autonomous navigation and obstacle avoidance enables the robot to perform tasks with reduced dependence on continuous human intervention [10]-[15].

In this work, an AI-driven voice recognition and autonomous control system for mobile robots is developed. The proposed system processes spoken commands using a machine learning-based recognition model and converts the recognized speech into robotic control actions. The robot combines voice-based commands with autonomous navigation and obstacle detection to achieve reliable movement in dynamic environments. The performance of the proposed system is evaluated using voice recognition accuracy, response time, navigation accuracy, and obstacle detection accuracy [16]-[20].

The experimental results show that the proposed system achieves 99.2% voice recognition accuracy with a response time of 200 ms**,** while navigation and obstacle detection achieve 98.8% and 98.9% accuracy, respectively. These results demonstrate the feasibility of integrating machine learning-based speech recognition with autonomous robotic control. The study also investigates the effect of noisy environments on recognition performance and identifies noise robustness as an important area for future improvement.

2.Machine Learning-Based Voice Recognition And Data Analysis

The voice-controlled operation of the proposed solar-powered autonomous robot is enhanced using a machine-learning-based speech-recognition approach as shown in Figure 1. The system recognizes five commands: Forward, Backward, Left, Right, and Stop.

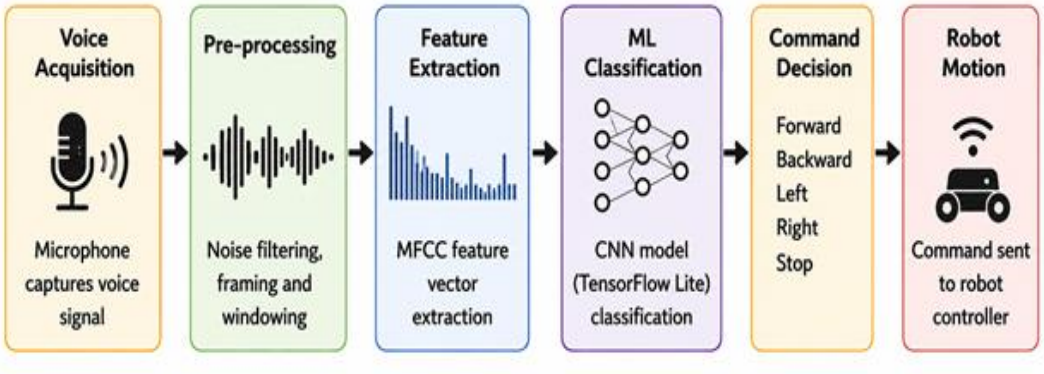


Figure 1.proposed methodology machine learning-based voice recognition and data analysis

The proposed methodology begins with voice acquisition, where a microphone captures the user's voice command. The captured signal undergoes pre-processing, including noise filtering, framing, and windowing. MFCC (Mel-Frequency Cepstral Coefficient) features are then extracted from the processed speech signal. These features are given to the CNN-based machine-learning classifier implemented using TensorFlow Lite to identify the command. The recognized command is classified as Forward, Backward, Left, Right, or Stop, and the corresponding command is sent to the robot controller to perform the required movement.

2.1 MFCC Feature Extraction

Pre-emphasis

$$y[n] = x[n] - \alpha x[n-1] \quad (1)$$

Framing and Windowing

$$x_w[n] = x[n]w[n] \quad (2)$$

Fast Fourier Transform (FFT)

$$X[k] = \sum_{n=0}^{N-1} x_w[n] e^{-j2\pi kn/N} \quad (3)$$

Power Spectrum

$$P[k] = \frac{1}{N}|X[k]|^2 \quad (4)$$

Mel Scale

$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad (5)$$

MFCC Coefficient

$$C_r = \sum_{k=1}^{K} \log(E_k) \cos\left[\frac{\pi r}{K}\left(k - \frac{1}{2}\right)\right] \quad (6)$$

MFCC feature extraction converts the speech signal into meaningful acoustic features for machine-learning classification. The process includes pre-emphasis, framing and windowing, FFT, power-spectrum calculation, Mel-scale conversion, and MFCC coefficient extraction**.** These steps capture the important frequency characteristics of the voice signal and produce a compact feature representation that can be given to the CNN classifier for recognizing the five voice commands.

2.2 Classification Model

Input Feature Vector:

$$X = [x_1, x_2, \ldots, x_m]^T \quad (7)$$

Classifier Output:

$$\hat{Y} = f(X; \theta) \quad (8)$$

Softmax Output:

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad (9)$$

Final Command:

$$\hat{Y} = \arg\max_i (P_i) \quad (10)$$

where

C = 5,representing Forward, Backward, Left, Right, and Stop.

The classification model uses the extracted MFCC features as an input feature vector and processes them through the machine-learning classifier. The Softmax function calculates the probability of each voice-command class, and the command with the highest probability is selected as the final prediction. The model classifies the speech into five commands: Forward, Backward, Left, Right, and Stop**.**

3.Results and discussion

3.1 Training and testing procedure:

Dataset	Purpose	Split Ratio
Training set	Model learning	70%
Validation set	Model tuning	15%
Testing set	Final evaluation	15%

The dataset is divided into three parts**:** 70% training data is used to train the machine-learning model, 15% validation data is used to tune and improve the model, and 15% testing data is used for final performance evaluation. This division helps assess how well the trained model performs on unseen data.

3.2 Performance metrics:

Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

F1-score:

$$F1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

Where: TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.
The performance of the machine-learning model is evaluated using accuracy, precision, recall, and F1-score. Accuracy measures the overall percentage of correctly classified samples, while precision indicates the proportion of correctly predicted positive samples. Recall measures the ability of the model to correctly identify actual positive samples. The F1-score provides a combined measure of precision and recall and is useful for evaluating the overall classification performance. Here, TP represents True Positive, TN represents True Negative, FP represents False Positive, and FN represents False Negative.

3.3 Confusion Matrix (Five-Class)

Actual / Predicted	Forward	Backward	Left	Right	Stop	Total
Forward	NFF	NFB	NFL	NFR	NFS	NF
Backward	NBF	NBB	NBL	NBR	NBS	NB
Left	NLF	NLB	NLL	NLR	NLS	NL
Right	NRF	NRB	NRL	NRR	NRS	NR
Stop	NSF	NSB	NSL	NSR	NSS	NS
Total	NF	NB	NL	NR	NS	N

The five-class confusion matrix evaluates how accurately the machine-learning model recognizes the five voice commands: Forward, Backward, Left, Right, and Stop. The rows represent the actual commands, while the columns represent the predicted commands. The diagonal values (NFF,NBB,NLL,NRR,NSS) indicate correctly recognized commands, whereas the off-diagonal values represent misclassifications. A higher number of correct predictions along the diagonal indicates better voice-recognition performance.

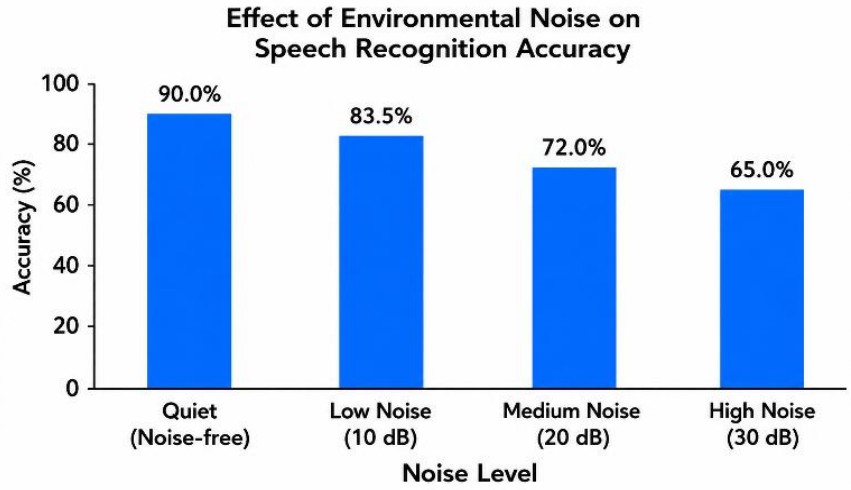
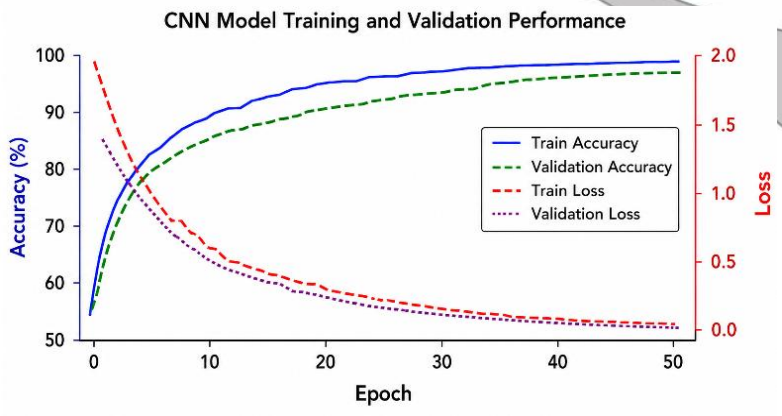


Figure 3. graph shows that speech-recognition accuracy

The graph shows that speech-recognition accuracy decreases as environmental noise increases. Accuracy is 90% in quiet conditions, decreasing to 83.5% at 10 dB, 72% at 20 dB, and 65% at 30 dB. This indicates that higher noise levels significantly affect voice recognition performance and highlights the need for noise-robust preprocessing and training.



Figue 4. graph shows that the CNN model

The graph shows that the CNN model learns effectively as the number of epochs increases. Training and validation accuracy steadily increase and reach approximately 99% and 97%, respectively, while both training and validation loss decrease toward zero. The small gap between training and validation accuracy indicates good model generalization with limited overfitting.

4. Conclusion

The proposed AI-driven voice recognition and autonomous control system for mobile robots successfully integrates machine learning-based speech recognition with autonomous navigation and obstacle detection to enable intelligent and real-time human–robot interaction. The experimental results demonstrate a voice recognition accuracy of 99.2% with an average response time of approximately 200 ms, indicating reliable and fast interpretation of spoken commands. The autonomous navigation and obstacle detection modules achieved accuracies of 98.8% and 98.9**%**, respectively, demonstrating effective movement control and safe obstacle avoidance.The system also maintained better recognition performance under different noise conditions compared with the conventional approach. However, recognition accuracy decreased as environmental noise increased, highlighting the importance of noise-robust preprocessing and training. Overall, the results confirm that the integration of artificial intelligence, voice recognition, and autonomous robotic control can provide an effective solution for mobile robot applications. Future work will focus on improving noise robustness, supporting a larger vocabulary and multiple languages, incorporating advanced deep learning models, and testing the system in complex real-world environments.

References

[1] M. Ayache, H. Kanaan, K. Kassir, and Y. Kassir, “Speech command recognition using deep learning,” in Proc. 6th Int. Conf. Advances in Biomedical Engineering (ICABME), 2021, pp. 24–29, doi: 10.1109/ICABME53305.2021.9604862.

[2] M. Soltanian, J. Malik, J. Raitoharju, A. Iosifidis, S. Kiranyaz, and M. Gabbouj, “Speech command recognition in computationally constrained environments with a quadratic self-organized operational layer,” in Proc. Int. Joint Conf. Neural Networks (IJCNN), 2021, pp. 1–6, doi: 10.1109/IJCNN52387.2021.9534232.

[3] Z. Lu and R. Huang, “Autonomous mobile robot navigation in uncertain dynamic environments based on deep reinforcement learning,” in Proc. IEEE Int. Conf. Real-time Computing and Robotics (RCAR), 2021, pp. 423–428, doi: 10.1109/RCAR52367.2021.9517635.

[4] R. Adamini, N. Antonini, A. Borboni, S. Medici, C. Nuzzi, R. Pagani, A. Pezzaioli, and C. Tonola, “User-friendly human-robot interaction based on voice commands and visual systems,” in Proc. 24th Int. Conf. Mechatronics Technology (ICMT), 2021, doi: 10.1109/ICMT53429.2021.9687192.

[5] A. Amano, H. Isshiki, and Y. Kato, “Autonomous mobile robot navigation by reinforcement learning considering pedestrian movement tendencies,” in Proc. 47th Annual Conf. IEEE Industrial Electronics Society (IECON), 2021, doi: 10.1109/IECON48115.2021.9589076.

[6] A. Misbullah, L. Farsiah, N. Nazaruddin, and F. Hermawan, “Voice-Zikr: A speech recognition system implementation for hands-free Zikr based on deep learning,” in Proc. IEEE Int. Conf. Cybernetics and Computational Intelligence (CyberneticsCom), 2022, doi: 10.1109/CyberneticsCom55287.2022.9865318.

[7] H. R. Vanamala, A. S. M. Akash, V. A., and S. Kumar, “Gesture and voice controlled robot for industrial applications,” in Proc. Int. Conf. Advancement in Technology (ICONAT), 2022, doi: 10.1109/ICONAT53423.2022.9725903.

[8] Z. Zhang, R. Li, Y. Wang, F. Dong, and K. Chen, “Research on voice interaction robot system based on the hybrid chat model of retrieval and generation,” in Proc. 2nd Int. Conf. Computing, Communication, Perception and Quantum Technology (CCPQT), 2023, doi: 10.1109/CCPQT60491.2023.00037.
[9] A. Patil, R. Khatavkar, S. Kardile, P. Kakade, and R. Pawar, “Computer controlled letter writing robotic arm using speech recognition,” in Proc. 7th Int. Conf. Computing, Communication, Control and Automation (ICCUBEA), 2023, doi: 10.1109/ICCUBEA58933.2023.10392038.
[10] L. Yu, Z. Wu, Y. Jiang, M. H. Tan, et al., “Autonomous navigation and obstacle avoidance for mobile robots in a tax environment,” in Proc. IEEE 13th Int. Conf. CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), 2023, doi: 10.1109/CYBER59472.2023.10256541.
[11] R. Tsuruta and K. Morioka, “Autonomous navigation of a mobile robot with a monocular camera using deep reinforcement learning and semantic image segmentation,” in Proc. IEEE/SICE Int. Symp. System Integration (SII), 2024, doi: 10.1109/SII58957.2024.10417188.
[12] R. J. Alitappeh, N. Mahmoudi, M. R. Jafari, and A. Foladi, “Autonomous robot navigation: Deep learning approaches for line following and obstacle avoidance,” in Proc. 20th CSI Int. Symp. Artificial Intelligence and Signal Processing (AISP), 2024, doi: 10.1109/AISP61396.2024.10475313.
[13] L. Jia, M. Yu, and Z. Zhang, “Robot visual perception and autonomous obstacle avoidance based on deep learning,” in Proc. 4th Int. Conf. Neural Networks, Information and Communication (NNICE), 2024, doi: 10.1109/NNICE61279.2024.10498312.
[14] S. Goetzee, K. Mihhailov, R. Van De Laar, K. Baraka, and K. V. Hindriks, “Audio-visual speech recognition for human-robot interaction: A feasibility study,” in Proc. 33rd IEEE Int. Conf. Robot and Human Interactive Communication (RO-MAN), 2024, pp. 930–935, doi: 10.1109/RO-MAN60168.2024.10731139.
[15] S. Nayagam, R. Sr., C. Sureshkumar, R. Kiruthiga, J. Akshya, M. D. Choudhry, and M. Sundarrajan, “IoT-enhanced path planning and obstacle avoidance for autonomous mobile robots,” in Proc. 3rd Int. Conf. Advancement in Technology (ICONAT), 2024, pp. 1–6, doi: 10.1109/ICONAT61936.2024.10774737.
[16] A. Muthukrishnan, S. Pushparani, N. Poongavanam, M. Arun, et al., “Real-time deep learning for dynamic scene understanding and autonomous robot navigation,” in Proc. Int. Conf. Cybernation and Computation (CYBERCOM), 2024, doi: 10.1109/CYBERCOM63683.2024.10803236.
[17] L. E. Rodríguez Raygoza, J. C. Tudon Martínez, J. de J. Lozoya-Santos, and L. C. Félix-Herrán, “Work in progress: Enhancing human-robot interaction through a speech and command recognition system for a service robot using ROS Melodic,” in Proc. 14th Int. Conf. Information Science and Technology (ICIST), 2024, doi: 10.1109/ICIST63249.2024.10805259.
[18] T. Veeranna, K. Pavan Kumar, M. Aruna, A. Verma, and T. Kurian, “Enhancing human-robot interaction through natural language processing and speech recognition techniques,” in Proc. 2nd Int. Conf. Computational and Characterization Techniques in Engineering & Sciences (IC3TES), 2024, doi: 10.1109/IC3TES62412.2024.10877529.
[19] A. Keefe and B. Hament, “Artificial intelligence (AI) voice module for robotic service dog,” in Proc. Systems and Information Engineering Design Symposium (SIEDS), 2024, doi: 10.1109/SIEDS61124.2024.10534692.
[20] S. Shrestha, Y. Zha, S. Banagiri, G. Gao, Y. Aloimonos, and C. Fermüller, “NatSGLD: A dataset with speech, gesture, logic, and demonstration for robot learning in natural human-robot interaction,” in Proc. 20th ACM/IEEE Int. Conf. Human-Robot Interaction (HRI), 2025, doi: 10.1109/HRI61500.2025.10974248.