Iran Agricultural Research (2026) 45(AI Special Issue) 55-63

Shiraz University

Iran Agricultural Research

Journal homepage: https://iar.shirazu.ac.ir

Research Article

# Magnetic microrobot navigation in a two-dimensional environment for gene delivery using natural language processing

**Khashayar Zare[a], Soleiman Hosseinpour[a*], Rojina Alizadeh[b], Dorsa Saeekia[b]**

[a] Department of Agricultural Machinery, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, I. R. Iran
[b] Alborz Province Education and Training Office for Brilliant Talents, Karaj, I. R. Iran

**ARTICLE INFO**

*Keywords:*
Gene delivery
Magnetic microbot
Microbotics
Natural language processing
Navigation and control

**Received**: 17 October 2025
**Revised:** 14 May 2026
**Accepted:** 18 May 2026

**ABSTRACT-** In this study, an interactive system combining the Vosk speech-to-text (STT) engine with a natural language processing (NLP) model was developed for controlling a disk-shaped magnetic microrobot. The system was designed based on the "human-in-the-loop" (HITL) approach, in which a human operator guides the microrobot in a two-dimensional space using voice commands. For performance analysis, a pre-trained DistilBERT model was employed, achieving an accuracy of 100% during training and 95.9% on validation data. In the test phase, the STT toolkit alone correctly converted only 32.65% of voice commands into text; however, after processing by the NLP model, the overall system accuracy increased to 79.59%. This indicates that the NLP model can compensate for errors in speech-to-text conversion. Furthermore, the developed model achieved an average processing time of 0.108 seconds, which is significantly faster than DistilBERT's 6.30 seconds, while maintaining classification accuracy. The results demonstrate that integrating an STT engine with a lightweight, accurate NLP model is an effective approach for controlling microrobot through voice commands. This system could pave the way for the development of voice-based navigation technologies in gene delivery, biological, and industrial applications.

## NOMENCLATURE

| Symbol | Definition | Unit |
|---|---|---|
| N | Number of turns of the Helmholtz coil | – (dimensionless) |
| I | Current flowing through the coil | Ampere (A) |
| R | Radius of the Helmholtz coil (distance from the center to each turn) | meter (m) |
| B | Magnetic flux density (magnetic field strength) generated by the coil | Tesla (T) |
| $\mu_o$ | Permeability of free space | $4\pi\times10^{-7}$ ($H\cdot m^{-1}$) |

## INTRODUCTION

Over the past few years, advances in microtechnology and nanotechnology have gradually introduced new possibilities in many scientific fields. One notable result of this progress is the development of microrobots. These tiny devices have drawn increasing attention since they can be used in various fields, such as medicine, environment, agriculture, and even the food industry (Sitti, 2017). Microrobots are particularly interesting because of their size and design. Their small scale allows them to reach places and carry out tasks that regular robots cannot reach, which has opened the view to applications that were once considered out of reach (Agrahari et al., 2020).

In the medical field, microrobots have provided innovative capabilities for the treatment of diseases. These tools can perform accurate procedures within the human body, such as targeted drug delivery, tissue sample collection, and even repair damaged tissues (Choi et al., 2021). For example, magnetic microrobots stimulated by an external magnetic field can navigate complex body environments and perform their tasks (Wang et al., 2020). In the food industry, microrobots have been used to separate harmful microorganisms, remove bacteria from food products, and produce value-added products, such as removing yeast cells from beverages (Villa et al., 2020).

In the environmental sector, microrobots have demonstrated significant potential, particularly in addressing water and soil pollution issues. These tools

*Corresponding Author: Associate Professor, Department of Agricultural Machinery, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, I. R. Iran
E-mail address: shosseinpour@ut.ac.ir
DOI: https://doi.org/10.22099/iar.2026.54562.1720

effectively clean environments contaminated with microplastics, remove heavy metals from water, and purify microbial contaminants (Parmar et al., 2018). To address the growing concerns regarding microplastic pollution in drinking water and food chains, the use of microrobots as a sustainable solution offers hope for alleviating this global issue (Zhou et al., 2021). One of the most critical features of microrobots is their ability to access specific, hard-to-reach areas with high precision (Hou et al., 2025), particularly at the cellular scale. These characteristics make them suitable for targeted drug or gene delivery in micro- or nano-dimensions (Schmidt et al., 2020; Qin et al., 2025). Consequently, microrobots are shaping the future of microbiology, medicine, and nanotechnology-based therapies (Chen et al., 2025).

The most essential challenge in the use of microrobots is their accurate control and navigation in complex environments (Shen et al., 2023). Some limitations, such as the inability to unify hardware components, such as sensors and actuators, make the control of microrobot movement more difficult (Medina-Sánchez and Schmidt, 2017). Magnetic microrobots, stimulated by an external magnetic field, are a suitable option for overcoming the challenges of propulsion and navigation at this scale. These microrobots can move precisely and perform their tasks in various environments, from biofluids to contaminated aqueous environments (Sitti, 2017).

In some methods, microrobot navigation is typically based on complex dynamical modeling and simplified assumptions, such as uniform magnetization or system linearity (Rahmer et al., 2017). Although these models aid in improved control, they have limitations when facing the complex environments and dynamics of microsystems. Furthermore, classical feedback systems are often inefficient in adapting to sudden environmental changes (Xu et al., 2015).

The advancement of artificial intelligence, especially methods based on Reinforcement Learning (RL) (Sutton and Barto, 1998), has opened up new possibilities for addressing the challenges involved in controlling microrobots. Salehi et al. (2024) explored this method using fixed magnetic actuation in the z-direction with Helmholtz coils and two permanent magnets, each attached to a servomotor for controlling the motion in the x- and y-planes to navigate a disk-shaped magnetic microrobot in a controlled laboratory setting. In their experiment, the microrobot control system was trained using the Soft Actor-Critic (SAC) and TRPO algorithms, which enabled it to gradually learn how to behave under different servomotor angles and inputs to navigate the microrobot to its target point in a two-dimensional plane (Salehi et al., 2024).

Another branch of artificial intelligence is Natural Language Processing (NLP) (Chowdhary, 2020). NLP has also received increasing attention in robotics. NLP can create a communication bridge between human intent and robotic execution (Park et al., 2024). For robotic systems to perform appropriate actions in various scenarios, it is essential that they possess comprehensive knowledge of the physical principles of their environment and the ability to interact effectively with users. Additionally, robots typically perform well in predictable environments with fixed task parameters; however, in unpredictable environments, the cost of designing specific control rules and specialized tools can limit their flexibility and adaptability. NLP can serve as a crucial tool to enhance Human-Robot Interaction (HRI) (Patidar and Soni, 2024), enabling robots to interpret, process, and respond to natural language commands ( Obaigbena et al., 2024; Vanzo et al., 2020). This enhanced communication between humans and machines allows for more effective interaction with robots and provides better control and guidance in task execution (Kahuttanaseth et al., 2018).

In this study, an integrated system based on the Vosk Speech-to-Text (STT) engine and an NLP model for controlling a disk-shaped magnetic microrobot is presented. This system utilizes a human-in-the-loop (HITL) approach. In this approach, a human user observes the workspace and guides the microrobot using voice commands in two-dimensional space. This system facilitates simpler and more accurate guidance of a microrobot to the target position, and the user can navigate the microrobot without the need for dynamic calculations or environmental simulations. Furthermore, this represents an effective step toward better human-machine interaction and microrobot navigation for gene delivery. Although gene delivery was not experimentally implemented in this study, precise and reliable microrobot navigation at micro-scale dimensions is a fundamental prerequisite for targeted biomedical applications, including gene delivery. The proposed framework establishes a controllable and stable navigation infrastructure that can be extended in future studies to integrate payload-carrying mechanisms and controlled release systems. Therefore, this study represents a foundational step toward the development of NLP-assisted magnetic microrobots for targeted gene delivery applications.

## MATERIALS AND METHODS

### *Microrobot and magnetic actuation system*

The present study utilized a disk-shaped magnetic microrobot fabricated from a neodymium (NdFeB) alloy, grade N45, with a triple-layer Nickel and Cobalt coating and axial magnetization (Webcraft Company, Gottmadingen, Germany). This microrobot has a radius of 750 µm and a thickness of 500 µm. Additionally, the magnetic actuation system designed by Salehi et al. (2024) was used to control the position. The actuation system comprises a pair of Helmholtz coils and two permanent magnets connected to two separate servomotors (Fig. 1).

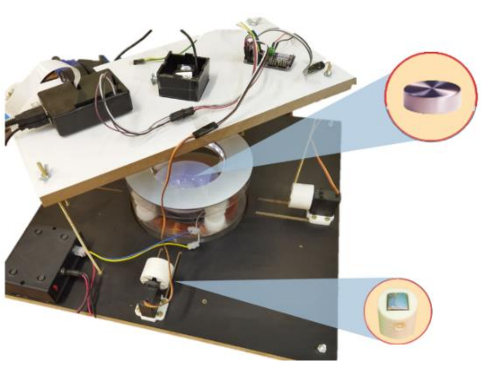

**Fig. 1.** The microrobot actuation system including a magnetic microrobot, two permanent magnets, and a Helmholtz coils.

The servomotors rotate the permanent magnets to generate changes in the magnetic field and gradient. The permanent magnets used in the system were cubic with dimensions of 20 × 20 × 3 $mm^3$, made from a neodymium alloy grade N42 with a triple-layer Nickel and Cobalt coating. The selected servomotors were MG996R model (Tower Pro Pte Ltd., Singapore) with an operational angle of 0° to 180°, placed 16 cm from the center of the workspace. This configuration allows the servomotors to change the angles of the permanent magnets, thereby adjusting the magnetic field and its gradient, which, in turn, directs the motion of the microrobot along the X and Y axes.

Helmholtz coils with a diameter of 15 cm were constructed using copper wire with a diameter of 1.3 mm and 166 turns, wound around a transparent Plexiglas bobbin. These coils were used to generate a uniform magnetic field that set the orientation of the magnetic moment of the microrobot along the Z-axis and simultaneously caused the microrobot to float on the water surface during the experiments (Yanmin et al., 2019).

Based on the Biot–Savart law, the magnetic field strength produced by the single-axis Helmholtz coils used to generate a force to float the microrobot on the surface of the water is a function of the number of turns (N), current passing through the coils (I), and radius (R). This relationship is expressed by the following equation (Hurtado-Velasco and Gonzalez-Llorente, 2016):

$$B_Z(Z)\big|_{|Z|\leq\frac{a}{2}} \approx \frac{8\mu_0 NI}{5\sqrt{5}a} \qquad \text{Eq. (1)}$$

Considering this configuration and utilizing Eq. (1), the single-axis Helmholtz coils can produce a uniform magnetic field with a maximum intensity of 6.9 millitesla (mT) at the center of the workspace (Yousefi and Nejat Pishkenari, 2021).

Then a 6-cm Petri dish was placed at the center of the Helmholtz coils and servomotors to ensure that the microrobot remained within the area where the magnetic field was uniform. The distance between the servomotors and Petri dish was fixed at 16 cm. After setting up the system, the microrobot was placed into a dish filled with deionized water to prevent air bubbles from forming, as they could disrupt the microrobot’s movement or make it difficult for the image processing part to detect the microrobot during the experiment.

In addition, the rotation of the servomotors was controlled using a Raspberry Pi 3B V1.2 (Raspberry Pi Foundation, Cambridge, UK). The servomotors were connected to the Raspberry Pi via a 16-channel PWM driver module with 12-bit accuracy and a PCA9685 chip (Adafruit Industries, New York, USA). A 5 V power supply was used to energize the PCA9685 chip, which controlled the servomotors. Ultimately, the two-dimensional workspace limited the movement of the microrobot to the XY plane. Fig. 1 shows the configuration of the magnetic actuation system.

*Microrobot detection*

Image processing techniques were employed to determine the position of the microrobot in the XY plane. Detection was performed using a Logitech QuickCam Pro 5000 USB camera (Logitech International S.A., Lausanne, Switzerland) with a resolution of 1920 × 1080 pixels and a frame rate of 30 fps. To regulate the ambient light intensity and ensure optimal visibility for accurate microrobot recognition, a 15-watt LED strip and 3000-watt dimmer were used.

Image processing was implemented using the *OpenCV* library in Python version 3.11.7. First, the latest frame captured by the camera was cropped to the workspace and then color processing was applied. To enhance efficiency and analysis speed, the image was converted to grayscale so that pixel values were mapped within a single-color band (0-255). Subsequently, additional refinements were applied to the binary image using two primary morphological operations, *erosion* and *dilation*. These processes helped in identifying boundary lines of objects in the frame. Ultimately, the most prominent object was detected, and the largest contour was used as the basis for accurate microrobot localization (Fig. 2).

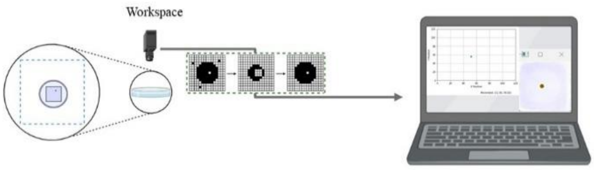


**Fig. 2.** Image processing using erosion and dilation to detect microrobot in the workspace and display it and its position on the laptop screen.

*Speech to text engine*

In this study, an STT engine was used to convert spoken voice inputs into textual word sequences. During the speech recognition process, a list of possible text outputs was generated, from which the most relevant option corresponding to the original audio signal was chosen. For this purpose, the Vosk toolkit was used in combination with the Python *speech-recognition* package, which activated the microphone and processed the user's voice input.

Vosk is an open-source speech processing tool based on the Kaldi framework that operates locally and does not require an online server connection. This local capability enables speech recognition to be performed entirely on the user device. Vosk is recognized for its ease of use and high accuracy. The model “*vosk-model-small-en-us-0.15*,” trained on diverse datasets, was utilized to convert spoken language into text. Although its accuracy is lower than that of cloud-based engines, it offers faster processing because it does not require data transmission to remote servers and remains fully functional without an internet connection.

*Natural language processing*

NLP models are typically optimized for processing structured textual data and deliver high levels of accuracy and performance in such contexts. Despite textual inaccuracies, NLP models can effectively interpret and classify texts. Therefore, spoken commands were first converted to text and then processed using NLP techniques. Details regarding the dataset used for the NLP implementation are provided in the Dataset and Language Model Implementation sections.

*Dataset*

The first and most critical step in developing a language model is data collection and the creation of high-quality datasets. For this study, a list of English words and phrases corresponding to four main directional concepts–"Downward", "Forward", "Left", and "Right"–was compiled. These terms are commonly used by both lay users and specialists alike. In addition, irrelevant, ambiguous, or unclear words and phrases were categorized separately under the label "Not recognized".

Table 1 shows the number of samples (words and phrases) in each category. The selected words and phrases were used to train the language model to improve its accuracy in interpreting and processing the directional commands.

**Table 1.** The number of each direction in the data set

| Direction | Number of data |
|---|---|
| Downward | 68 |
| Forward | 92 |
| Left | 104 |
| Right | 94 |
| Not recognized | 136 |

To develop the language model, text selection, preprocessing, and training were performed to enable the precise interpretation of user commands for microrobot control. To prevent the need for the installation of different Python packages or powerful hardware, the model training was performed in Google Colab using Python version 3.10.12. The dataset was uploaded to Google Drive to simplify the preprocessing and model development in the Colab environment. The uploaded file was loaded using the pandas library. The dataset was split into training (80%), validation (10%), and test (10%) sets for training and validating the newly trained model. For use in TensorFlow, the tokenized data were converted into a dataset object to enable equal data processing and improve efficiency.

*DistilBERT NLP model*

DistilBERT (Sanh et al., 2019) is a lightweight and optimized version of the BERT model developed by Hugging Face. While preserving performance comparable to BERT (Koroteev, 2021), it is more computationally efficient and requires fewer resources than BERT. The next component is the word embedding layer. Numeric tokens produced by the tokenizer were converted into high-dimensional vectors (768 dimensions), representing the semantic and syntactic features of the words in the vector space. This model also works better with small datasets than the original BERT model.

The process involved data preprocessing, tokenization, model training, evaluation, and saving the trained models. DistilBERT tokenization was used to convert the input text into numerical sequences (tensors), which is essential for transformer-based models (Jaderberg et al., 2015). During tokenization, sentences exceeding 32 tokens were truncated, and shorter sentences were padded to 32 tokens. The transformer layers in DistilBERT consist of six layers, each composed of two main components: The first is multi-head attention, which allows the model to simultaneously examine the relationships between words from multiple perspectives. The second is the feedforward neural network, which integrates and transforms the information processed by the attention mechanism (Fig. 3).

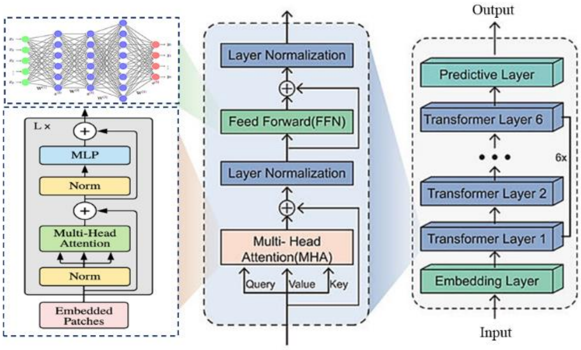


**Fig. 3.** The structure and operation of the DistilBERT model.

Finally, the transformer outputs were passed to a fully connected classification layer. In this study, the model was trained for five epochs on the training data using the Adam ( Kingma & Ba, 2015) optimizer with a learning rate of 0.00004 and a loss parameter of *SparseCategoricalCrossentropy*. A validation set was used to prevent overfitting and assess performance. Upon completion of the training process, the model and tokenizer were saved for the microrobot navigation.

*Rotation angles of the servomotors*

All models, including image processing, STT, and the developed NLP model, were executed on an Asus ROG Strix 512LI laptop (ASUSTeK Computer Inc., Taipei, Taiwan). The laptop was equipped with an Intel Core i7-10750H processor (Intel Corporation, Santa Clara, USA), an Nvidia GeForce GTX 1650Ti graphics card (Nvidia Corporation, Santa Clara, USA) with 4 GB of dedicated memory and 16 GB of RAM. To improve user voice clarity during speech-to-text conversion, a Q Sent T20 desktop microphone (Q Sent, Shenzhen, China) with a frequency response of 30-16000 Hz and a sensitivity of 58 dB (Decibel) was used.

First, the accuracy of the STT tool was evaluated by having the user verbally input dataset words and phrases, followed by an assessment of the speech-to-text conversion. The NLP model was then evaluated using a test dataset. To assess the NLP performance before and after integration with the STT model, textual data were first entered directly into the NLP model, and its accuracy and processing speed were recorded. In the second stage, the integrated STT-NLP performance was evaluated using the same dataset.

The analyses demonstrated how integrating the STT model with the NLP model affected text recognition capabilities and improved overall system accuracy. For this purpose, a user verbally issued commands, and the performance of both the STT and NLP models in guiding the microrobot was assessed. The "*vosk-model-small-en-us-0.15*" was selected because of its small size and high processing speed.

The position of the microrobot was tracked using an overhead camera, and its movement was analyzed using image-processing techniques. During operation, the user spoke different directional commands, which were first converted to text using the STT model. After undergoing preprocessing similar to the model training stage, the text was fed into the language model. If no valid command was detected or if the spoken instruction was unclear, a message appeared on the monitor indicating that the system was unable to interpret the input. In such cases, no new command was sent to the actuation unit, and its state remained the same. Tokenization was applied to the input text to enhance speed and recognition accuracy. Whenever the tokenized form failed to match any of the predefined commands, the system categorized the input as "Not recognized."

After receiving the spoken commands from the human operator, the language model processed the input and resulted in specific rotation angles for the permanent magnets in the clockwise or counterclockwise directions. Servo motor 0 controlled the vertical microrobot movement; specifically, the clockwise rotation moved the microrobot upward, and the counterclockwise rotation moved it downward. Servo motor 1 controlled the horizontal movement, which means that clockwise rotation moved it to the left, and counterclockwise rotation moved it to the right. In the case of input errors (operator commands), no changes were applied to the servo angles.

The rotation angles were transmitted to a Raspberry Pi via a local Wi-Fi network using the Python socket communication protocol. The Raspberry Pi then sent the data to the PCA9685 driver module via I2C protocol. The PCA9685 module forwarded the rotation angles to the corresponding servomotors. This configuration established a HITL control system that directly combined human voice command input into the control mechanism to boost decision accuracy and unite operational performance (Fig. 4).

## RESULTS AND DISCUSSION

### *Pre-trained distilBERT model performance*

The model was trained over five epochs, and the resulting performance metrics are reported in Table 2. The model achieved a strong performance in terms of accuracy and reduced loss.

The training progress of the model is illustrated in Fig. 5. The graph demonstrates a stable and improving trend from the beginning to the end of the fifth epoch. The absence of significant fluctuations in the loss and accuracy values indicated effective training and minimal overfitting.

When evaluated independently, without integration with the STT model, the proposed model achieved a perfect accuracy of 1.000 and a very low loss value of 0.0378. These results demonstrate that the model correctly classified all textual inputs and was effectively trained with respect to both the syntactic structure and semantic representation. The perfect accuracy highlights the model's strong capability to interpret and classify textual data within the targeted application domain.

The confusion matrix corresponding to the model is presented in Fig. 6. As shown, no misclassifications occurred across any of the five classes, and the model correctly identified and classified all the test data. This confirms the high accuracy of the model in processing linguistic data within the scope of this experiment.

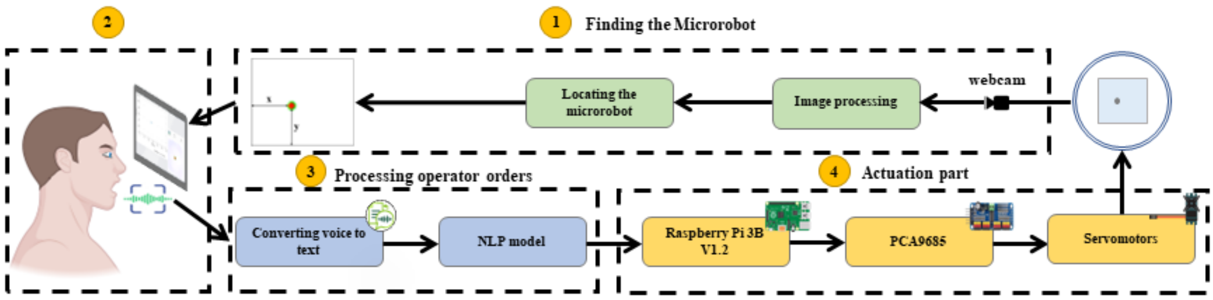


**Fig. 4.** human-in-the-loop (HITL) navigation framework for microrobot control: (1) A webcam captures top-view images of the workspace, (2) the operator observes the microrobot's position and issues a movement command, (3) the speech to text (STT) model converts the spoken command into the text, which is then processed by the natural language processing (NLP) model, and (4) the generated control command is transmitted from the laptop to the Raspberry Pi using socket, which relays the signal to a PCA module to actuate the servomotors. The servomotors rotate to the specified angles, enabling the microrobot to move in the desired direction.

**Table 2.** Parameters of the saved model after five epochs

| Parameter | Value |
|---|---|
| Training loss | 0.047 |
| Training accuracy | 1.000 |
| Validation loss | 0.134 |
| Validation accuracy | 0.959 |

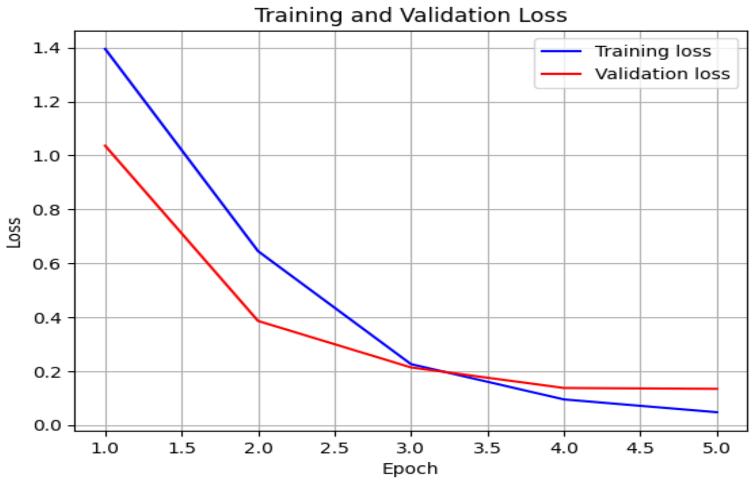


**Fig. 5.** Training and validation loss over five epochs.

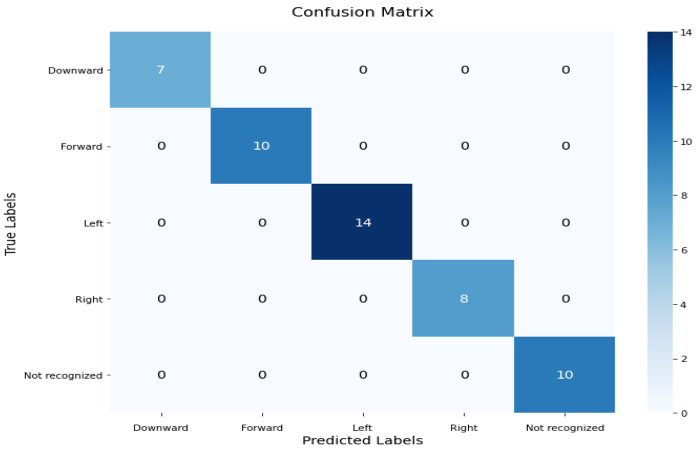


**Fig. 6.** Confusion matrix of the natural language processing (NLP) model on the test dataset.

*NLP model integrated with the STT system*

To measure the performance of the developed STT-NLP hybrid system, a subset of test data that was distinct from the training data was used. Initially, the STT toolkit correctly converted only 16 out of 49 spoken inputs into text and achieved an accuracy of 32.65%. As they were fed to the NLP model and processed by it, the number of correctly classified commands increased to 39, and the overall system accuracy improved to 79.59%. A key observation is that because the NLP model was designed to interpret meaning rather than rely on perfect transcription, low STT inaccuracies had minimal impact on the final output. (Table 3).

In cases where the STT toolkit misrecognized certain words or spoken phrases by the human operator, the NLP model was still able to extract the intended meaning and command the microrobot correctly. This proves the robustness of the NLP model in handling noisy or partially incorrect input and confirms that its combination with the STT toolkit provides a reliable solution for the accurate interpretation of spoken instructions (Table 4).

Additionally, as shown in the confusion matrix in Fig. 7, the STT-NLP combination successfully classified unrelated and invalid inputs in the "Not recognized" class. This feature prevents incorrect commands from being executed and prevents unwanted microrobot movement (Fig. 7).

As shown in Table 4, the average time taken to convert the operator's voice to text by the STT model was 0.017 s. Meanwhile, the average time taken by the NLP model to classify and send commands to the Raspberry Pi to apply the desired settings to the magnetic actuator system (servomotors) was 6.3 s. However, all commands that were classified correctly resulted in the magnetic actuator system, causing the microrobot to move along the desired path as intended by the operator (Fig. 8).

**Table 3.** speech to text (STT) performance when evaluated using the DistilBERT model

| Direction | Total samples | Correct transcriptions | Text accuracy (%) |
|---|---|---|---|
| Downward | 7 | 4 | 57.14 |
| Forward | 10 | 3 | 33.33 |
| Left | 14 | 1 | 7.14 |
| Right | 8 | 4 | 50.00 |
| Not-recognized | 10 | 4 | 40.00 |
| Overall | 49 | 16 | 32.65 |

**Table 4.** Accuracy and processing time of the speech to text - natural language processing (STT-NLP) model for each direction in the test dataset

| Direction | Number of samples | STT-NLP accuracy (%) | STT toolkit time (s) | STT-NLP time (s) |
|---|---|---|---|---|
| Downward | 7 | 71.42 | 0.016 | 6.38 |
| Forward | 10 | 60.00 | 0.015 | 6.28 |
| Left | 14 | 71.42 | 0.020 | 6.19 |
| Right | 8 | 100.00 | 0.021 | 6.28 |
| Not-recognized | 10 | 100.00 | 0.013 | 6.31 |
| Overall | 49 | 79.59 | 0.017 | 6.30 |

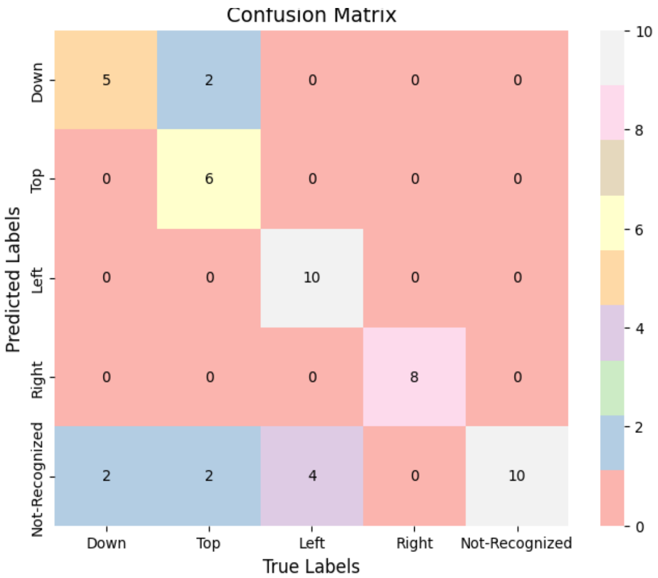


**Fig. 7.** Confusion matrix of the speech to text - natural language processing (STT-NLP) model for each direction in the test dataset.

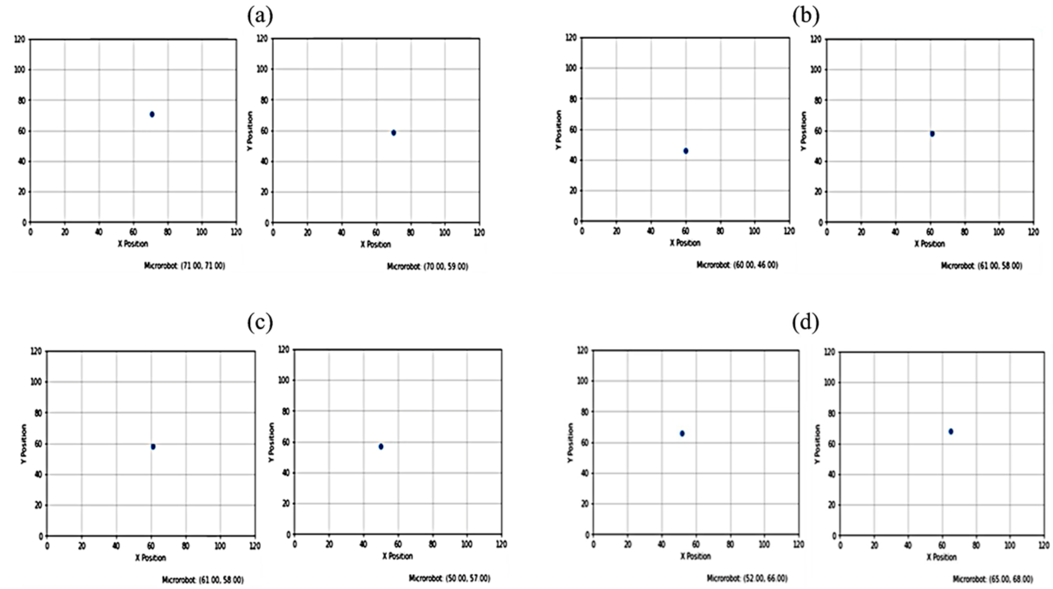


**Fig. 8.** The microrobot's position is visible at the initial position and after the operator's command in the following directions: (a) down, (b) forward, (c) left, and (d) right.

The control strategies for microrobots have evolved from direct physical interfaces to sophisticated, autonomous systems. Lu et al. (2019) introduced a human-microrobot interface based on acoustic frequency modulation, where commands were manually input through typing or music. While innovative, this method requires users to learn a new form of communication and lacks an error correction mechanism. In contrast, Li et al. (2017) and Salehi et al. (2024) shifted toward full autonomy by employing vision-based path planning and deep reinforcement learning (DRL), respectively. These approaches have successfully eliminated the need for real-time human intervention, achieving remarkable success rates in dynamic environments. However, they also removed the human operator from the decision-making loop, which presents a critical limitation in sensitive applications, such as medical procedures, where human judgment remains essential. The system proposed in this study bridges this gap by introducing a voice-based NLP interface that keeps the "HITL" while making the interaction both intelligent and robust.

The results further highlight the advantages of this approach. Although Salehi et al. (2024) achieved a 100% success rate in autonomous navigation and Li et al. (2017) demonstrated adaptive path planning, neither addressed the fundamental challenge of intuitive human-robot communication. The direct frequency control proposed by Lu et al. (2019) offers an immediate response but suffers from the absence of error handling. Despite the initial STT engine's poor accuracy of only 32.65%, our system leveraged NLP to compensate for these errors, achieving an overall accuracy of 79.59% with a processing time of only 0.108 s while successfully navigating the microrobot in the intended direction with complete reliability. This demonstrates that intelligent language processing can transform an imperfect input channel into a robust, control interface. Unlike the black-box learning of reinforcement learning or the rigid mapping of acoustic signals, our method provides a transparent, intuitive, and error-tolerant interaction paradigm that makes it particularly suitable for clinical and biological applications, where both precision and human oversight are paramount.

## CONCLUSION

While current research in microrobot navigation primarily focuses on complete automation with minimal human intervention, this study successfully validated the hypothesis that integrating a pretrained NLP DistilBERT model with the Vosk STT engine can create a robust and intuitive HITL control interface for magnetic microrobots. The proposed system effectively bridges the gap between direct human commands and autonomous machine execution, addressing the limitations of previous approaches. Unlike physical interfaces that require new communication methods or fully autonomous systems that exclude human oversight, this voice-based NLP interface enables intelligent and error-tolerant interactions. The results confirm that while the STT engine alone is insufficient for reliable control, the NLP model compensates for its inaccuracies by interpreting the semantic meaning and filtering invalid inputs, thereby transforming an imperfect channel into a dependable command interface. This capability, demonstrated by the system's ability to achieve reliable navigation and reject unrecognized commands, establishes the proposed method as a promising foundation for developing safer microrobot navigation systems. By enabling precise control through natural voice commands, this interface lays the groundwork for sensitive applications, such as targeted gene delivery, in which an operator can supervise and guide a gene-coated microrobot to a specific location with high accuracy, ensuring both safety and adaptability in dynamic physiological environments.

## FUNDING

There is no funding available for the current research.

## CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Conceptualization: Khashayar Zare and Soleiman Hosseinpour; Methodology: Khashayar Zare; Software: Khashayar Zare; Validation: Khashayar Zare and Soleiman Hosseinpour; Formal analysis: Soleiman Hosseinpour; Investigation: Soleiman Hosseinpour; Resources: Soleiman Hosseinpour; Data curation: Khashayar zare; Writing—original draft preparation: Khashayar Zare; Writing—review and editing: Khashayar Zare, Rojina Alizadeh, and Dorsa Saeekia; Visualization: Khashayar Zare; Supervision: Soleiman Hosseinpour; Project administration: Soleiman Hosseinpour.

## DECLARATION OF COMPETING INTEREST

The authors declare no conflicts of interest.

## ETHICAL STATEMENT

Approval was granted by the Ethics Committee of University of Tehran.

## DATA AVAILABILITY

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

## ACKNOWLEDGMENTS

The authors thank Dr. Amar Salehi for his invaluable help in designing and building the device. We also gratefully acknowledge the Department of Agricultural Machinery, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, for providing the essential support required for this project.

## REFERENCES

Agrahari, V., Agrahari, V., Chou, M. L., Chew, C. H., Noll, J., & Burnouf, T. (2020). Intelligent micro-/nanorobots as drug and cell carrier devices for biomedical therapeutic advancement: Promising development opportunities and translational challenges. *Biomaterials, 260,* 120163. https://doi.org/10.1016/j.biomaterials.2020.120163

Chen, X., Li, Y., Wang, C., Chen, Z., Xu, Z., Xia, F., Yan, Y., & Gao, M. (2025). Micro/nanorobots in antimicrobial therapy: Addressing challenges of antibiotic resistance. *Materials Today Bio*, *32*, 101936. https://doi.org/10.1016/j.mtbio.2025.101936

Chowdhary, K. (2020). *Natural language processing. Fundamentals of Artificial Intelligence.* Springer natuer link.

Hou, C., Yang, S., Chan, K. F., & Zhang, L. (2025).

Micro/nanorobots with remote sensing capabilities for intelligent and miniaturized biomedical platforms. *Biomedical Instrumentation, 1*, 100024. https://doi.org/10.1016/j.bmi.2025.100024

Hurtado-Velasco, R., & Gonzalez-Llorente, J. (2016). Simulation of the magnetic field generated by square shape Helmholtz coils. *Applied Mathematical Modelling, 40*(23-24), 9835-9847. https://doi.org/10.1016/j.apm.2016.06.027

Jaderberg, M., Simonyan, K., & Zisserman, A. (2015). *Spatial transformer networks. Advances in Neural Information Processing Systems, 28*, 2017-2025.

Kahuttanaseth, W., Dressler, A., & Netramai, C. (2018). Commanding mobile robot movement based on natural language processing with RNN encoderdecoder. In 2018 5th International Conference on Business and Industrial Research (ICBIR) (pp. 161-166). IEEE. https://doi.org/10.1109/ICBIR.2018.8391185

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv, 1412*, 6980. https://doi.org/10.48550/arXiv.1412.6980

Koroteev, M. V. (2021). BERT: A review of applications in natural language processing and understanding. *arXiv*, *2103*, 11943. https://doi.org/10.48550/arXiv.2103.11943

Li, T., Chang, X., Wu, Z., Li, J., Shao, G., Deng, X., Qiu, J., Guo, B., Zhang, G., He, Q., Li, L., & Wang, J. (2017). Autonomous collision-free navigation of microvehicles in complex and dynamically changing environments. *ACS Nano*, *11*(9), 9268-9275. https://doi.org/10.1021/acsnano.7b04525

Lu, X., Zhao, K., Liu, W., Yang, D., Shen, H., Peng, H., Guo, X., Li, J., & Wang, J. (2019). A human microrobot interface based on acoustic manipulation. *ACS Nano*, *13*(10), 11443-11452. https://doi.org/10.1021/acsnano.9b04930

Medina-Sánchez, M., & Schmidt, O. G. (2017). Medical microbots need better imaging and control. *Nature*, *545*(7655), 406-408. https://doi.org/10.1038/545406a

Obaigbena, A., Lottu, O. A., Ugwuanyi, E. D., Jacks, B. S., Sodiya, E. O., & Daraojimba, O. D. (2024). AI and human-robot interaction: A review of recent advances and challenges. *GSC Advanced Research and Reviews*, *18*(2), 321-330. https://doi.org/10.30574/gscarr.2024.18.2.0070

Park, S., Wang, X., Menassa, C. C., Kamat, V. R., & Chai, J. Y. (2024). Natural language instructions for intuitive human interaction with robotic assistants in field construction work. *Automation in Construction, 161,* 105345. https://doi.org/10.1016/j.matdes.2023.111735

Patidar, R., & Soni, S. (2024). Enhancing human-robot interaction through advanced natural language processing techniques*. International Journal of Scientific Research in Engineering and Management (IJSREM), 08*, IJSREM35803. https://doi.org/10.55041/IJSREM35803

Qin, X., Xu, R., Wu, J., Liu, Y., Wang, T., Tu, H., Li, J., & Pang, Z. (2025). Recent advances in engineering nano/microrobots for tumor treatment. *Acta Pharmaceutica Sinica B. 6222*, 6252. https://doi.org/10.1016/j.apsb.2025.10.003

Rahmer, J., Stehning, C., & Gleich, B. (2017). Spatially selective remote magnetic actuation of identical helical micromachines. *Science Robotics*, *2*(3), eaal2845. https://doi.org/10.1126/scirobotics.aal2845

Salehi, A., Hosseinpour, S., Tabatabaei, N., Soltani Firouz, M., & Yu, T. (2024). Intelligent navigation of a magnetic microrobot with model-free deep reinforcement learning in a real-world environment. *Micromachines*, *15*(1), 112. https://doi.org/10.3390/mi15010112

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv*, *1910*, 01108. https://doi.org/10.48550/arXiv.1910.01108

Schmidt, C. K., Medina-Sánchez, M., Edmondson, R. J., & Schmidt, O. G. (2020). Engineering microrobots for targeted cancer therapies from a medical perspective. *Nature Communications, 11*(1), 5618. https://doi.org/10.1038/s41467-020-19322-7

Shen, H., Cai, S., Wang, Z., Ge, Z., & Yang, W. (2023). Magnetically driven microrobots: Recent progress and future development. *Materials & Design, 227*, 111735. https://doi.org/10.1016/j.matdes.2023.111735

Sitti, M. (2017). *Mobile microrobotics*, 271. Retrived from: https://www.researchgate.net/publication/319173891_MOBILE_MICROROBOTICS

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge: MIT press.

Vanzo, A., Croce, D., Bastianelli, E., Basili, R., & Nardi, D. (2020). Grounded language interpretation of robotic commands through structured learning. *Artificial Intelligence*, *278*, 103181. https://doi.org/10.1016/j.artint.2019.103181

Wang, Q., Li, T., Fang, D., Li, X., Fang, L., Wang, X., Mao, C., Wang, F., & Wan, M. (2020). Micromotor for removal/detection of blood copper ion. *Microchemical Journal, 158*, 105125. https://doi.org/10.1016/j.microc.2020.105125

Xu, T., Yu, J., Yan, X., Choi, H., & Zhang, L. (2015). Magnetic actuation based motion control for microrobots: An overview. *Micromachines*, *6*(9), 1346-1364. https://doi.org/10.3390/mi6091346

Yanmin, F., Dixiao, C., Yuguo, D., De, G., & Lin, F. (2019). The design and control of magnetized cell-based microrobot for targeting drug delivery. In 2019 IEEE 14th International Conference on Nano/Micro Engineered and Molecular Systems (NEMS) (pp. 273-276), Ieee. https://doi.org/10.1109/NEMS.2019.8915608

Yousefi, M., & Nejat Pishkenari, H. (2021). Independent position control of two identical magnetic microrobots in a plane using rotating permanent magnets. *Journal of Micro-Bio Robotics, 17*(1), 59-67. https://doi.org/10.1007/s12213-021-00143-w

Zhou, A., Zhang, Y., Xie, S., Chen, Y., Li, X., Wang, J., & Zou, J. (2021). Microplastics and their potential effects on the aquaculture systems: A critical review. *Reviews in Aquaculture, 13*(1), 719-733. https://doi.org/10.1111/raq.12496.